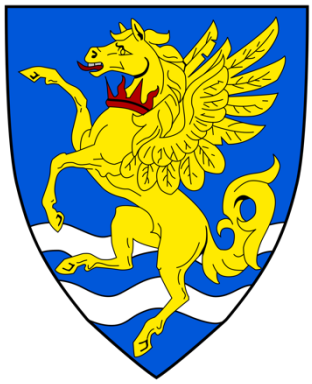
Glycation of Extracellular Matrix: A Solid-State NMR Study

Rui Li



Robinson College and Department of Chemistry
University of Cambridge

September 2019

This thesis is submitted for the degree of Doctor of Philosophy.

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Degree Committee.

Rui Li

September 2019

Glycation of Extracellular Matrix: A Solid-State NMR Study

Rui Li

The main aim of this thesis is to investigate the glycation products, especially glycation intermediate products, generated in the glycation reaction model in vitro extracellular matrix (ECM) and ribose-5-phosphate using solid-state nuclear magnetic resonance (SSNMR) spectroscopy. The background and motivations of this project are outlined in chapter 1.

Chapter 2 reviews the current understanding of the composition and organisation of the ECM, the structure of collagen and the critical roles collagen plays in the ECM. Chapter 3 introduces the glycation reaction and summarises the glycation products identified to date, with a discussion of the most widely-used characterisation techniques and in vitro model systems for studying the glycation reaction and glycation products. Chapter 4 covers the basic theory of SSNMR spectroscopy, which is the major analytical technique used in this thesis, and the principles behind the SSNMR experiments used in this project. Above three chapters provide the essential background knowledge for studying the glycation reaction chemistry and the glycation products in in vitro collagenous ECM model using SSNMR spectroscopy.

Protocols for preparing, glycating and characterising in vitro ECM samples as well as all relevant experimental details are included in chapter 5, which is followed by a discussion in chapter 6 on the development of a method to extract collagen from in vitro ECM. In chapters 7 and 8, glycation products identified in different in vitro glycation models using SSNMR are outlined, where chapter 7 focuses on glycation products derived from the glycation reaction between non-isotope-enriched collagenous materials and ^{13}C-enriched glycating agents, while chapter 8 concentrates on the effects of glycation on lysine and arginine residues in collagenous samples using ^{13}C, ^{15}N-enriched collagenous materials and ^{13}C-enriched glycating agents. Chapter 9 discusses the development of a protocol to extract cellular metabolites, specifically carbonyl metabolites, and estimate their glycation potential, in order to understand the contribution that cell necrosis may have on glycation in patients. Conclusion and future work are outlined in chapter 10.

This page is dedicated to my parents for their unreserved and unconditional love and support.

“父兮生我，母兮鞠我。

拊我畜我，长我育我。

顾我复我，出入腹我。”

Acknowledgements

Nothing can really be accomplished solely on one's own – my PhD work is no exception. So many people have helped me in different ways during this difficult yet rewarding journey.

Foremost, I would like to thank all the members of the Duer group, past and present: Dave, Ying, Veronica, Yang, Rakesh, Ieva, Ulya, Anna, Adrian and Thomas for their training, support, generous assistance and good company throughout my time in Cambridge. It would not be possible to forget all the laughter we had in our cosy office 251, all the botanical garden trips we wandered, all the pastries and bread we shared from that lovely French bakery and homemade goodies from Adrian and Thomas (we are so well fed!). You guys are such lovely PhD buddies! Particularly, I would like to thank Prof. Melinda Duer for giving me the opportunity to join the group, for being such a supportive and enthusiastic supervisor during my time here in Cambridge and for giving me the encouragement and freedom to pursue my ideas. I could never ask for more than what I have got from a PhD supervisor. Special thanks go to Dr. Wing Ying Chow for her guidance through the exploration of the exciting DNP-SSNMR world, for answering all my stupid questions patiently and for recommending places to go and restaurants to try during my visit in Berlin.

Heartfelt thanks go to Dr. Peter Grice, Mr. Duncan Howe, Mr Andrew Mason for training me on solution-state NMR spectrometers and running a smooth NMR service in the department. Thank Dr. Karin Muller for all her help with electron microscopy and sharing her wisdom during our sessions. I will miss your cooking very much!

I would like to acknowledge all the funding bodies for their financial support throughout my PhD: Cambridge Trust and China Scholarship Council for my PhD scholarship, without which I could never have the opportunity to pursue a PhD in Cambridge; iNEXT for financing the trips to Berlin and the access to DNP-SSNMR and 700 MHz SSNMR spectrometers; Robinson College and the Department of Chemistry for their financial contribution towards my academic conference expenditures.

Last but definitely not the least, I thank all my family and friends near or far for loving me, supporting me and keeping me company at my best and my worst, without whom I could never have made it this far. Especially Jo Xinyu, you have lit up my life in Cambridge, and I feel tremendously lucky to have you.

List of Abbreviations

Amino Acid	3-letter	1-letter
Arginine	Arg	R
Cysteine	Cys	C
Glycine	Gly	G
Histidine	His	H
Hydroxylysine	Hyl	-
Hydroxyproline	Hyp	O
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Tryptophan	Trp	W
Tyrosine	Tyr	Y

Acronym	Definition
3-DG	3-Deoxyglucosone
3DG-H	3-Deoxyglucosone Hydroimidazolone
AAA	Amino Acid Analysis
AEC	3-Amino-9-Ethylcarbazole
AGEs	Advanced Glycation Endproducts
BSA	Bovine Serum Albumin
C7	Sevenfold Symmetric Phase Shift Scheme
CSA	Chemical Shift Anisotropy
CEL	N^{ε}-(carboxylethyl)lysine
CL	Cell Lysate
CMA	N-(carboxymethyl)arginine
CMC	S-(carboxylmethyl)cysteine
CMhL	N^{ε}-(carboxylmethyl)hydroxylysine
CML	N^{ε}-(carboxylmethyl)lysine
COL	Collagenous
CP	Cross Polarisation
DARR	Dipolar-Assisted Rotational Resonance
DCP	Double Cross Polarisation
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribose Nucleic Acid
ECM	Extracellular Matrix
ELN	Elastin
F6P	Fructose-6-Phosphate
FN	Fibronectin
Fru	Fructose
FSOBs	Foetal Sheep Osteoblasts
FTL	Freeze-Thaw Lysis
Fuc	Fucose
G6P	Glucose-6-Phosphate
GA	Glyceraldehyde
GAG	Glycosaminoglycan
G-H	Glyoxal hydroimidazolone

Glc	Glucose
GOLD	Glyoxal Lysine Dimer
HETCOR	Heteronuclear Correlation
HSA	Human Serum Albumin
LNL	Liquid Nitrogen Lysis
KSPG	Keratan Sulfate Proteoglycan
MAS	Magic Angle Spinning
MG-H	MethylGlyoxal-derived 5-Hydro-5-imidazolone
MODIC	MethylglOxal Dimer Imidazole Crosslink
MOLD	MethylglOxal Lysine Dimer
MS	Mass Spectroscopy
MW	Molecular Weight
NC	Non-Collagenous
NMR	Nuclear Magnetic Resonance
ns	Number of Scans
PC7	See POST-C7
PDSD	Proton-Driven Spin Diffusion
PG	Proteoglycan
PLL	Poly-L-Lysine
POST-C7	Permutationally Offset Stabilised C7
PTMs	Post-Translational Modifications
R5P	Ribose-5-Phosphate
Rib	Ribose
S/N	Signal to Noise
SEM	Scanning Electron Microscopy
SPC-5	Supercycled POST-C5
SQ-DQ	Single Quantum-Double Quantum
SSNMR	Solid-State Nuclear Magnetic Resonance
TEDOR	Transferred Echo Double Resonance
TEM	Transmission Electron Microscopy
THP	Tetrahydropyrimidine
VSMCs	Vascular Smooth Muscle Cells

Contents

1 Introduction

Diabetes is an increasingly serious global problem due to the constantly high sugar level circulating in the body, causing numerous diabetic complications and ultimately affecting all organs and tissues. Though the pathology of diabetes and diabetic complications is complex and not fully understood, one of the factors is proposed to be a dramatic increase in the frequency of glycation and accumulation of advanced glycation end-products (AGEs). Glycation is a non-enzymatic chemical reaction between aldehyde groups of reducing sugars and protein amine groups, via a cascade of molecular rearrangements, oxidations, and crosslinking reactions, ultimately forming AGEs.

The most likely candidates to be affected by glycation are long-lived proteins, such as collagen in the extracellular matrix (ECM) where the cellular protection mechanisms against glycation are not present. Glycation and AGE formation degrade ECM structure and mechanical strength of the tissue and disturb critical cell behaviours, including signalling, adhesion, proliferation and differentiation, eventually causing widespread cell necrosis. Cell necrosis, an inherent part of many diseases, potentially releases reactive metabolites into the ECM, prompting further ECM glycation and damage. Other consequences of glycation include the production of more reactive carbonyl species and reactive oxygen species, resulting in a vicious circle in which reactive molecules are produced in the ECM, triggering further ECM glycation and more deleterious effects.

The glycation reactions are complex with many possible reaction routes, and their detailed mechanisms still remain poorly understood even after decades of research. Although some AGEs have been well-characterised in literature, increasing evidence suggests that it is not AGEs but glycation intermediate products that are primarily responsible for the structural and functional ECM changes caused by glycation.

This project aims to uncover the structures of abundant glycation products, including both intermediate and end products, present at the complex chemical equilibrium of the glycation reaction in the ECM, and exploring their effects on collagen molecular structure and potentially

fibrillar organisation using a combination of solid-state Nuclear Magnetic Resonance (SSNMR) spectroscopy and multiple analytical techniques, including Transmission Electron Microscopy (TEM), fluorescence spectroscopy and biological assays. SSNMR spectroscopy can identify specific AGEs and probe the effects of glycation on the ECM structure and molecular dynamics down to the atomic level. We use vascular smooth muscle cell and osteoblast in vitro ECM as experimental models to mimic the pathological condition of diabetic blood vessels and bone, where diabetes pathology has particular deleterious effects, such as increased bone fragility and risk of fracture and increased risk of cardiovascular diseases. Our primary focus is the effect of glycation on collagen. Collagen fibrils constitute the dominant component of most ECM in the body and are essential for ECM organisation and cell support. We employ isotope labelling schemes in a range of selectively ^{13}C- and ^{15}N-enriched ECM and ^{13}C-enriched biogenic glycating agents, U-$^{13}C_5$-ribose and U-$^{13}C_5$-ribose-5-phosphate, in order to overcome the sensitivity issue of SSNMR and use SSNMR spectroscopy to identify glycation products in the ECM and study the effects of glycation on the ECM. This approach allows us to look at the effects of glycation on specific amino acids, probe the glycation sites in collagen and identify the major glycation products. Such knowledge will advance approaches to inhibit the consequences of diabetic glycation and oxidative damage, and ultimately lead to more effective treatments for diabetes and diabetic complications.

Primary glycation sites are on lysine and arginine, which are not only critical to balance charges on the collagen fibril surface but also essential to maintain normal cell function, integrin binding and cell recognition. Modifications by glycation on these residues is already known to result in abnormal cell behaviours, though the mechanism of disruption is not yet clear. Stiffer ECM in diabetic patients, normal ageing and even cancer, has been attributed to collagen crosslinks from glycation. However, our results here show that glycation crosslinks are relatively insignificant and unlikely to make a huge difference to ECM properties, while the major glycation products that cause all the pathological consequences of glycation are lysine and arginine sidechain modifications. We propose that instead of crosslink formation, the disruption of charge balance on the surface of collagen fibrils and the addition of bulky AGE substituents to collagen lysine and arginine sidechains makes a considerable contribution to the increases in collagen stiffness and other property changes.

2 Collagen and The Extracellular Matrix (ECM)

The ECM is an essential non-cellular element of all tissues, such as skin, bone, blood vessels, etc. It fills the extracellular space and provides a physical scaffold for resident cells in almost every tissue. Additionally, the ECM involves in and regulates diverse cellular activities via cell-ECM interactions, which is essential to maintain tissue functions. The tissue-specific requirements on the ECM, such as structural, mechanical and biological properties, are achieved by its complex composition and organisation, which are under tight enzymatic regulation to maintain tissue homeostasis. Disrupted ECM composition or organisation potentially changes the ECM-cell interactions, affects tissue homeostasis and eventually leads to impaired tissue functions and diseases.

The ECM is composed of two structurally distinctive components, which are the basement membrane (the BM) and the interstitial matrix (the IM). The BM is a specialised form of the ECM which bridges between resident cells and the IM. The 3D architect of the ECM is built by hundreds of different types of proteins, polysaccharides and water in a tissue-specific manner. All the ECM components collaboratively achieve the intact ECM structure and functions.

This chapter reviews essential ECM composition and functions with a special focus on collagen because collagen, as the major ECM component, plays diverse and essential roles to maintain tissue structural integrity and overall tissue homeostasis and guarantee the specialised tissue functions. Alterations in the collagen structure can lead to disrupted ECM composition or organisation and impaired cell-ECM interactions, ultimately causing a variety of diseases and tissue dysfunctions.

Most information included in this chapter is from *Cell Biology of Extracellular Matrix* [1], if not specifically cited.

2.1 Collagen

In mammals, the ECM is composed of around 300 proteins with different sizes and functions. The most prevalent ECM protein is collagen, which is found in many tissue ECMs, such as skin, bone, blood vessels, cartilage, etc. For example, in bone ECM, collagen can make up ~90% its dry weight [2]. This ubiquitous distribution of collagen suggests that collagen is a versatile biomaterial that can match tissue-specific requirements, such as providing tissue scaffolds, supporting biological functions and meeting mechanical requirements.

This section reviews collagen structure and organisation and briefly covers the diverse roles collagen plays in the ECM and its relations to diseases, in order to provide key background knowledge for studying and understanding collagen structure and property changes under diseased conditions, i.e. glycation later in this thesis.

Unless specific cited in the text, information is summarised from the following books: *Collagen: Primer in Structure, Processing and Assembly* [3], *Collagen: Structure and Mechanics* [4] and *Fibrous Proteins: Structures and Mechanisms (Chapter 14)* [5].

2.1.1 The Collagen Family

Collagen is a family of structural proteins sharing a unique secondary structure: the triple helix. This distinctive triple helix is formed by an association of three left-handed polypeptide chains, called α chains, in a right-handed manner. Three α chains within one collagen molecule can be identical, forming a homotrimer, or different, forming a heterotrimer. Up to now, forty-two α chains have been identified, giving rise to a total of twenty-eight collagen types. [6][7][8]

Besides triple helical domains (or collagenous domains, COL domains), collagen molecules have non-collagenous domains (NC domains). Depending on the location, length and number of NC domains, the collagen family is categorised into subfamilies, such as fibrillar collagens, network-forming collagens, the fibril-associated collagens with interrupted triple helices (FACITs), membrane-associated collagens with interrupted triple helices (MACITs), multiple triple helix domains and interruptions (MULTIPLEXINs), etc. (Table 2-1).

Fibrillar collagens are the most abundant collagen types. They have the fewest number of NC domains per molecule, only occurring at either side of the COL domain, in the N- and C-termini (an example molecule is shown in Figure 2-1 A). Among different fibrillar collagens, this work focuses on collagen type I, the most widely occurring collagen found in skin, tendon, bone, cornea, lung and the vasculature. Major collagen types relevant to this project are detailed in Table 2-2.

Table 2-1 Twenty-eight types of collagens are numbered by Roman numerals from I to XXVIII. They can be divided into different categories based on their molecular and aggregation structures, including but not limited to fibrillar collagens, network collagens, the fibril-associated collagens with interrupted triple helices (FACITs), membrane-associated collagens with interrupted triple helices (MACITs) and multiple triple helix domains and interruptions (MULTIPLEXINs). [7]

Category	**Type**
Fibrillar	I, II, III, V, XI, XXIV, XXVII
Network	IV, VI, VIII, X
FACIT	IX, XII, XIV, XVI, XIX, XX, XXI, XXII, XXVI
MACIT	XIII, XVII, XXIII, XXV
MULTIPLEXIN	XV, XVIII
Others	VII, XXVIII

Table 2-2 Major collagen types mentioned in this project. α chains are numbered by Arabic numerals. Collagen type IV has two different molecular compositions. [9]

Type	**α Chain Composition**	**Structural Features**	**Tissues**
Fibrillar Collagens			
I	$[\alpha1(I)]_2\alpha2(I)$	300 nm long molecules, forming periodic fibrils	Skin, tendon, bone, etc.
III	$[\alpha1(III)]_3$	300 nm long molecules; often with type I	Skin, muscle, blood vessels, etc.
V	$\alpha1(V)\alpha2(V)\alpha3(V)$	390 nm long molecules with globular N-terminal domain; often with type I	Similar to type I; cell cultures, foetal tissues
Network Collagen			
IV	$[\alpha1(IV)]_2$, $\alpha2(IV)$ $\alpha3(IV)$, $\alpha4(IV)$, $\alpha5(IV)$	Two-dimensional parallelogram network	All basal laminas

2.1.2 Collagen Molecules

Before folding into a triple helix, α chains are transcribed and translated from specific collagen genes, and extensively post-translationally modified by enzymes, including hydroxylation of proline (P) and lysine (K) to hydroxyproline (O) and hydroxylysine (Hyl) respectively, and glycosylation of Hyl. Three α chains then associate to form a procollagen molecule with a central, long COL region and two NC regions on each side, called N- and C-propeptides. Propeptides are removed by enzymes, leaving short N- and C-telopeptides. The removal of the propeptide segments transforms procollagen into collagen (Figure 2-1 B). Typical collagen type I molecules are composed of three α chains with ~1000 amino acids each, giving a molecule of ~300 nm in length (L) and ~1.5 nm in diameter (d).

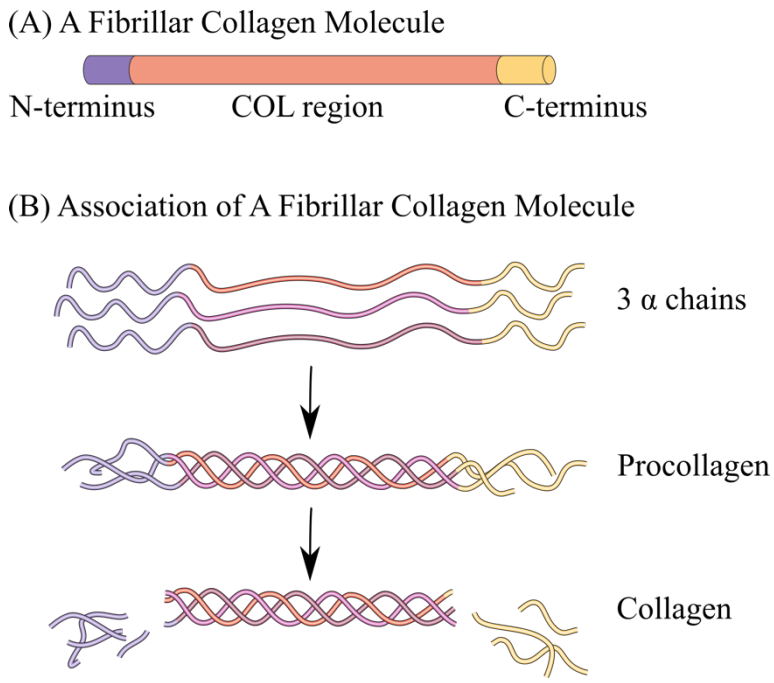


Figure 2-1 A: A fibrillar collagen molecule is composed of a uninterrupted central collagenous region (COL region) and NC regions, N-terminus and C-terminus, on each side of its COL region. B: Three α chains associate into a collagen molecule, via a procollagen. This figure is modified from Figure 1 in reference [10].

The Collagen Triple Helix: Structure and Stability

The collagen triple helix is stabilised by several factors. Each α chain has a conformation of three residues per turn, and every third residue in the triple helical region has to be the smallest

amino acid, glycine (G) [11]. This arrangement ensures G is always at the axial centre of a triple helix and its sidechain (-H) allows a close approach of the three α-chains and a tightly wound triple helix. It also makes a collagen molecule composite a multiple of G-X-Y repeating tripeptides (Figure 2-2). The X and Y positions can be occupied by any amino acids but are most often occupied by P (28%) and 4-hydroxyproline (O) (38%) respectively. Of all possible G-X-Y combinations, the most frequently occurring motif in collagen is GPO (10.5%) [7][12].

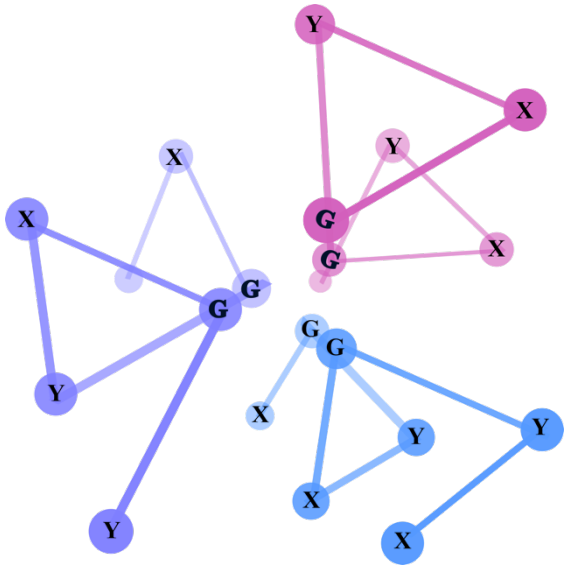


Figure 2-2 Collagen triple helix is composed of three α chains and each α chain has three residues per turn and is entirely composed of G-X-Y repeats to ensure G is always at the centre of a triple helix to allow tight packing.

P and O stabilise triple helices by their specific steric restrictions posed by their puckered rings. O at Y positions further stabilises the triple helix via a stereoelectronic effect. The stabilisation effects of P and O on the triple helix structure are well studied using collagen-like model peptides. [3][5][7][13] When the X and Y positions are not occupied by P and O, interchain salt bridges between two charged residues may also contribute to collagen triple helix stability. [14]

Additionally, the triple helix is stabilised by interchain hydrogen bonds primarily between the backbone N-H group of G and the backbone C=O group of the X position amino acid on the adjacent chain ($N\text{-}H_{(G)} \cdots O{=}C_{(X)}$) across the axis of the triple helix (Figure 2-3). Other hydrogen bonds are also proposed either mediated by water or bonded directly between residues; some of them are identified in the crystal structures of collagen-like peptides. For example, $C_{\alpha}\text{-}H_{(G/Y)} \cdots O{=}C_{(X/G)}$ or $N\text{-}H_{(X)} \cdots O{=}C_{(G)}$ or $N\text{-}H_{(X)} \cdots H2O \cdots O{=}C_{(G)}$ when the X

position is not occupied by P or O. Hydroxyl groups on O can also participate in hydrogen bonding. [3][5][15]

It is worth noting that the collagen triple helix pitch changes along the sequence. Imino acid-rich regions are more tightly wound than imino acid-dilute regions. Pure GPO molecules are closer to a 7/2 helix, while native collagen type I is closer to a 10/3 helix [16]. Also, local arrangements within collagen higher-order structures affect local triple helix conformation [16].

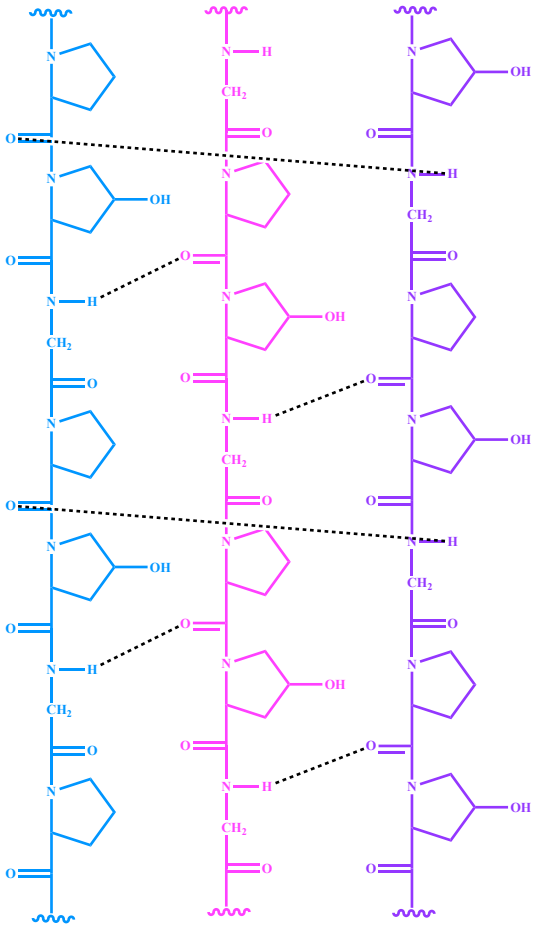


Figure 2-3 An illustration of collagen interchain hydrogen bonding, indicated by black dotted lines, between the N-H groups of G and C=O groups of the residue on X position, here P.

The Non-Collagenous (NC) Domains

NC domains are essential to collagen structure at both molecular and fibril level. They play an important role in triple helix folding and fibril formation [17]. They also include sites for intermolecular crosslinks to stabilise collagen higher-order structures [18][19].

2.1.3 Collagen Fibrils and Higher-Order Structures in the ECM

In the ECM, collagen molecules participate in higher-order assemblies in a complex and hierarchical manner to ultimately form tissue scaffolds and achieve biological functions. In terms of fibrillar collagens, the first step is to form collagen fibrils.

In collagen type I rich fibrils, five collagen molecules pack laterally with a D stagger between adjacent molecules [20] (Figure 2-4 top). Because L (the length of each collagen molecule) is not an integral multiple of D (in the tendon, D = ~ 67 nm = ~ 0.22 L), this packing will result in gap zones and overlap zones. Generally speaking, gap zones (~ 0.54 D) are longer than overlap zones (~ 0.46 D). These alternating gap- and overlap-zones along the fibril give the unique banding patterns seen in EM images.

While G is packed around the triple helix inner axis, sidechains of residues at X and Y positions tend to point outwards interacting with neighbouring molecules and these intermolecular interactions, electrostatic and hydrophobic interactions [21], are suggested to drive fibril formation and maintain fibril stability. Collagen fibrils are further stabilised by enzymatically-derived crosslinks between collagen molecules via K or Hyl residues in terminal NC regions, such as dehydro-hydroxylysino-norleucine (deH-HLNL) and dehydro-lysinonorleucine (deH-LNL). These divalent crosslinks can then undergo further reaction to form mature trivalent crosslinks, such as lysyl pyridinoline (L-Pyr) and hydroxylysyl pyridinoline (HL-Pyr) (Figure 2-4 bottom) [22]. Depending on the developmental stage of collagenous tissues, additional non-enzymatic crosslinks are also found, commonly derived from the glycation reaction (will discuss in Chapter 3). Unlike well-regulated enzymatic crosslinks, glycation-derived crosslinks are not regulated by enzymes, but formed via pure chemical reactions, and often associated with ageing and pathologies.

In the ECM, collagen fibrils are not comprised solely of one collagen type; they are co-assembled from different collagen types and NC proteins and eventually construct the different complex hierarchical structures of different tissues in order to adapt their tissue-specific requirements. For example, so-called type I collagen fibrils also contain a minor quantity of collagen type III in skin [23] and blood vessels [24]. The presence of major collagen types, such as collagen type I, and minor collagen types, such as collagen type III and V, in collagen fibrils in different tissues is relatively well-studied and their roles in different tissues are slowly

being revealed. Quantitatively minor collagen types are suggested to be critical to collagen fibril formation in initiating fibril genesis and regulating fibril diameters. For example, collagen type V and XI in tendon have been suggested to regulate fibril assembly in developing tendon [25].

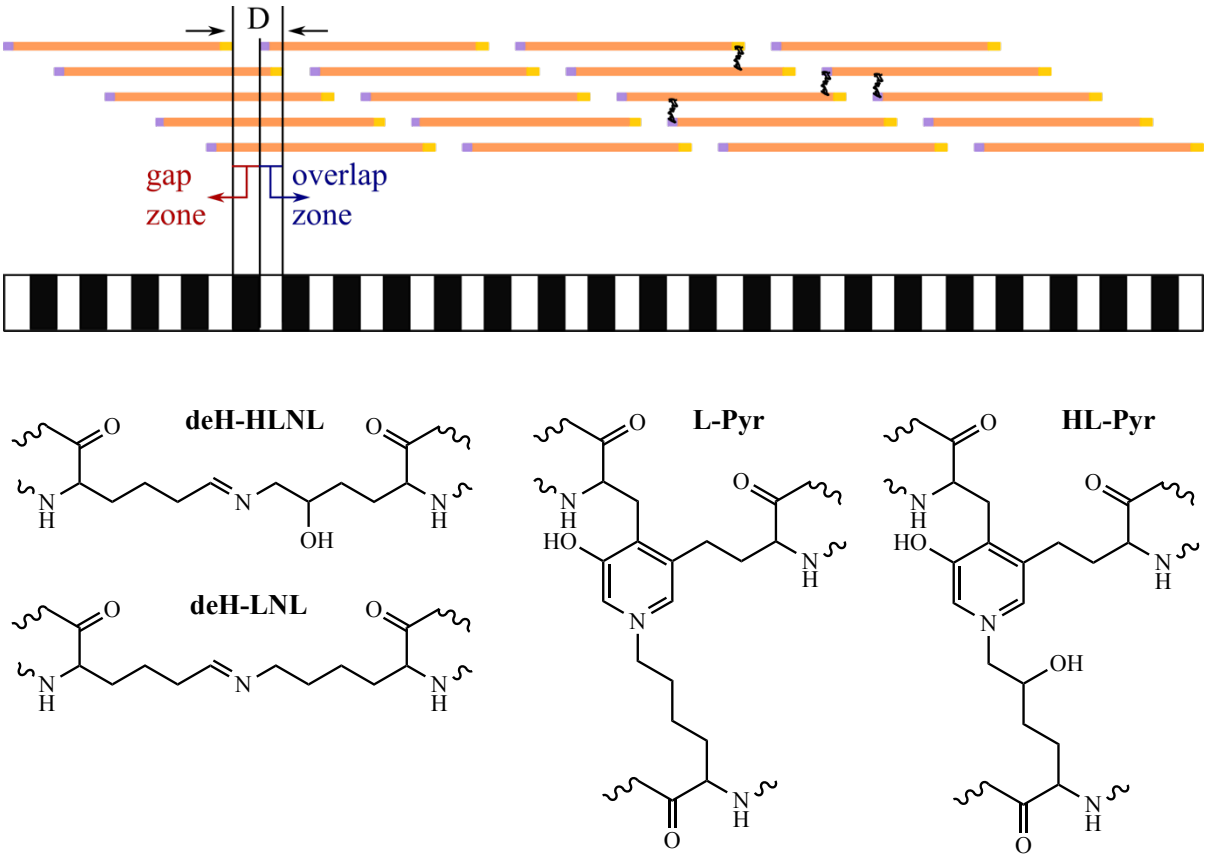


Figure 2-4 Top: Collagen molecules assemble with a D stagger between two molecules along their long axis and pack laterally with other four molecules into collagen fibrils. This molecular organisation within collagen fibrils results in gap zones and overlap zones and gives banded patterns in electron micrographs. Enzymatic crosslinks further stabilise collagen fibrils (indicated by the black wavy lines). Bottom: Chemical structures of example enzymatic crosslinks.

2.1.4 Roles of Collagen

The primary function of collagen is to provide mechanical stability required by different tissues. For example, in blood vessels, the collagen scaffold is assembled to resist multidirectional stresses, whereas in tendon or bone, the collagen superstructure is required to resist axial stress. The mechanical properties of collagen fibrils primarily are suggested to depend on the enzymatic crosslinks and their maturation and are also related to the minor collagen types co-assembled in the fibrils. Addition of glycation crosslinks may change the mechanical properties of collagen fibrils.

Collagen also have ligand binding sites for other ECM proteins and can achieve biological functions via collagen-cell surface receptor interactions. Collagen-protein and collagen-cell interactions play an important role in regulating ECM composition and structure as well as cell behaviours, such as adhesion, migration or proliferation. For example, GFOGER sequences on collagen are recognised by several integrins to transmit biological information between cells and their extracellular environments [26]. This means modifying some amino acid sites can easily affect cell-ECM communications.

Collagen, as well as many other ECM components, is not a static structure but is under dynamic remodelling. Collagen is constantly deposited, degraded and modified in the ECM, which is mediated strictly by different enzymes, such as matrix metalloproteinases (MMPs) and lysyl oxidase [27]. Modified collagen may also interrupt this regular remodelling, causing abnormal collagen accumulation.

2.2 The ECM, Collagen and Diseases

The ECM components and organisation are vital to both structural and functional homeostasis of tissues. However, abnormal ECM structure and function can result from both internal factors, such as genetic mutations and imbalanced enzymatic activities, and external factors, such as oxidative stress and non-enzymatic glycation reactions (see Chapter 3). Given its significance in regulating cell behaviours, any defect in the ECM can contribute to alterations of cellular pathways, eventually leading to tissue dysfunction and the development or progression of diseases.

A long list of diseases has been observed to associate with abnormal ECM structure or composition (either causing the disease or resulting from the disease), such as ageing, the development of diabetic complications, cancer genesis and progression. Consequently, the ECM can be treated as diagnostic and therapeutic targets for the above-mentioned diseases. Hence, a better understanding of both native ECM and diseased ECM will contribute to the development of novel therapies of ECM related diseases. [27][28]

Ageing, Diabetes and Diabetic Complications

Numerous changes have been observed in the ECM when tissues age, involving compositional, structural and mechanical changes. ECM changes associated with diabetes and diabetic complications are similar to those associated with ageing, suggesting common mechanisms behind those ECM changes. In this regard, diabetes is considered as accelerated ageing.

Increased stiffness has been observed in aged and diabetic tissues, which is widely attributed to the increase of crosslinking in aged tissues. The profile of enzymatic crosslinks changes with age from immature bivalent crosslinks to mature trivalent crosslinks, which is attributed to age-related mechanical property changes [22]. A second crosslinking origin recognised in aged tissues is glycation-derived crosslinks, a type of advanced glycation endproducts (AGEs) (see Chapter 3) [29][30]. However, whether the increase in crosslinks is the main reason behind age/diabetes-related tissue stiffening and which crosslink molecules contribute the most to the increased stiffness remain unrevealed.

Cancer

Cancerous tissues have been demonstrated to be stiffer than the surrounding normal tissues and the increased stiffness has been suggested to promote cancer progression in breast cancer via cancer cell-ECM interactions [31][32]. Increased collagen crosslinking and stiffening has been suggested to contribute to breast cancer invasion [33]. However, more work needs to be done to fully understand the molecular mechanism driving cancer ECM stiffening.

2.3 Non-Collagenous ECM Proteins

Other than collagen, various NC proteins, including proteoglycans (PGs), glycoproteins (GPs), etc., fill the space between cells and collagen scaffold in the ECM, providing connections

between collagen fibrils and between collagen fibrils and cells. Even though NC proteins are in minor quantities relative to collagen, their relative amount and composition of the NC proteins defines tissue features such as viscoelasticity. The structural modifications of NC proteins can also disrupt cell-ECM communications and lead to deleterious consequences. [34]

Minor proteins are present in lower amounts than collagens in in vivo tissues, thus their roles and their changes in diseases are much less documented compared to the major collagen types. In this project, the in vitro ECM used to mimic diseased ECM contains, in theory, most of the NC proteins - some might be lost during the decellularisation process – and therefore allows study of NC protein changes under disease conditions. A brief introduction to some NC proteins is included in this section to provide insights on possible changes related to ECM NC proteins in diseases.

Proteoglycans (PGs)

PGs are a group of proteins, consisting of a protein core with covalently-bonded glycosaminoglycan (GAG) chains. Hyaluronic acid (HA) is a special case in the PG family; it is not bound to a protein core but consists entirely of a long GAG chain. The most abundant ECM PGs are the small leucine-rich proteoglycans (SLRPs), which are characterised by relatively small protein cores containing leucine-rich repeats [35]. Because of the presence of GAG chains, the PG family members are negatively charged and extremely hydrophilic; therefore, they can interact with surrounding water molecules to maintain the hydration level of the ECM. GAGs with their surrounding water network are suggested to make a contribution towards resistance to compressive stress for tissues. [36][37][38]

Additionally, the PG family participates in a wide variety of activities, including regulating cell behaviours, collagen fibrillogenesis and ECM assembly. In the ECM, PGs interact with collagen via their protein cores and bridge between collagen fibrils to maintain the macrostructure of the ECM. [28][39]

Elastin (ELN)

ELN is critical to the elasticity and resilience of tissues, especially to tissues under repeated cycles of stretching and compressing such as blood vessels and the lung. In some aortic tissues, ELN can make up to 40% of dry, defatted tissue weight. [40]

In terms of amino acid composition, ELN is rich in hydrophobic and nonpolar amino acids. It has similar G and P molar ratios to collagen, ~33% and ~10% respectively and contains a small number of O. It is crosslinked similarly to collagen, via lysine derivatives. This amino acid composition and crosslinked nature make ELN insoluble and chemically inert.

ELN is the predominant protein to form elastic fibres that allow tissues to extend and recoil. Different elastin isoforms and elastic fibre morphologies from different assemblies are observed in different tissues to accommodate diverse tissue functions. [34][41][42]

Fibronectin (FN)

FN is a fibril-forming and cell-interacting ECM protein, which plays roles in mediating cell adhesion with the ECM and cell migration.

FN is extensively glycosylated and therefore, is categorised as a glycoproteins (GP). The GPs family describes glycosylated proteins with oligosaccharides covalently attached to specific amino acid residues, usually Asn, Ser or Thr. Around 200 different GPs have been identified in the ECM, covering a diverse range of structures and functions. They usually have multiple protein-binding domains to bind to cell surface receptors, collagen, PGs and other ECM components to realise the two-way communications between cells and ECM components. [1][43]

Other Remarks

A significant number of ECM components has been identified so far, including 28 types of collagen, more than 30 PGs, around 200 GPs, etc. Besides the major structural components reviewed above, many other proteins, such as growth factors and enzymes, and lipids are present in the ECM and essential to both the ECM and resident cells. [28][44] Within the

integral ECM, all components interact with one another and cells to maintain tissue integrity and homeostasis.

2.4 Summary

The ECM is a complex and dynamic structure that works as an important microenvironment for cells in almost all tissues. The complex ECM architecture is built by a sophisticated assembly of ~300 different ECM components, including water, collagen, proteoglycans, etc., and they are all under tight enzymatic regulations to maintain structural integrity and ensure tissue functions. Dysregulation can easily lead to abnormal ECM composition and structure and ultimately affect cell behaviours and tissue function.

The ECM is more than a physical scaffold for tissues. All the ECM components actively participate in mechanical and biological roles to complete full tissue functions. Constant and two-way cell-ECM communications exchange mechanical and biological information, which is critical to tissue homeostasis. Aberrant ECM can trigger different metabolic pathways and contribute to many diseases, such as diabetic complications and cancer.

Increases in tissue stiffness have been observed in aged, diabetic and cancerous tissue ECM, implying a similar pathological reason behind the stiffening in each case. One of the widely accepted reasons behind these pathologies is that increased glycation crosslinking alters mechanical properties, manipulates cell-ECM interactions and disrupts the regular collagen remodelling. However, direct evidence is still lacking to associate observed age-related changes to crosslinked collagen. Moreover, in vivo tissue ECM have non-collagenous components, which may also be affected by diseases.

3 The Glycation Reaction

The glycation reaction was first discovered and studied by food scientists as the browning effect. Only over recent decades, has glycation in biological systems been investigated intensively because of its implications in diseases, including ageing [45][46], diabetes and its complications [47][48][49][50][51][52], Alzheimer's disease [53][54][55][56] and cancer [57][58][59]. The pathologies of glycation are suspected to be very similar in all disease contexts, though as yet there has been no systematic study and the mechanisms behind the pathology are not fully understood. The belief is that the pathological species are advanced glycation endproducts (AGEs) formation and their accumulation in tissues.

This chapter includes the current understanding of the glycation reaction, reviews glycation products identified in different in vivo and in vitro systems, and evaluates commonly used characterisation techniques, in order to identify what is lacking in this research field and further advance the studies on the glycation chemistry and glycation products.

3.1 Glycation and Advanced Glycation Endproducts (AGEs)

Glycation is a non-enzymatic reaction between an amino group in proteins, DNAs and lipids, and a carbonyl group, such as reducing sugars, going through Maillard reaction, Amadori rearrangement and a cascade of subsequent reactions, eventually forming AGEs.

The whole process is extremely complicated with numerous possible routes, but a simplified scheme between a reducing sugar and a protein is shown in Figure 3-1. The glycation reaction is a pure chemistry reaction, i.e. not dependent on any enzymes, starting from nucleophilic attack of the amine group in the protein on the aldehyde group in the reducing sugar, forming, via a Schiff base, an Amadori product. Up to this point, the reaction is reversible. After formation of the Amadori product, a series of irreversible steps occur, including oxidation and dehydration, generating a complex group of glycation intermediates and eventually leading to the formation of AGEs.

Complications are introduced by several different possible reaction pathways during the glycation process, as indicated in Figure 3-1. Reducing sugar, Schiff base and Amadori

products can go through oxidation or degradation, forming more reactive (di-)carbonyl intermediates. The (di-)carbonyl intermediates (some examples are listed in Table 3-1) are much more reactive than the original reducing sugars and therefore can trigger more vigorous glycation reactions and generate more glycation products. Free radicals have also been detected during the glycation process, further increasing the variety of reaction routes and enlarging the pool of glycation products [60][61][62][63][64].

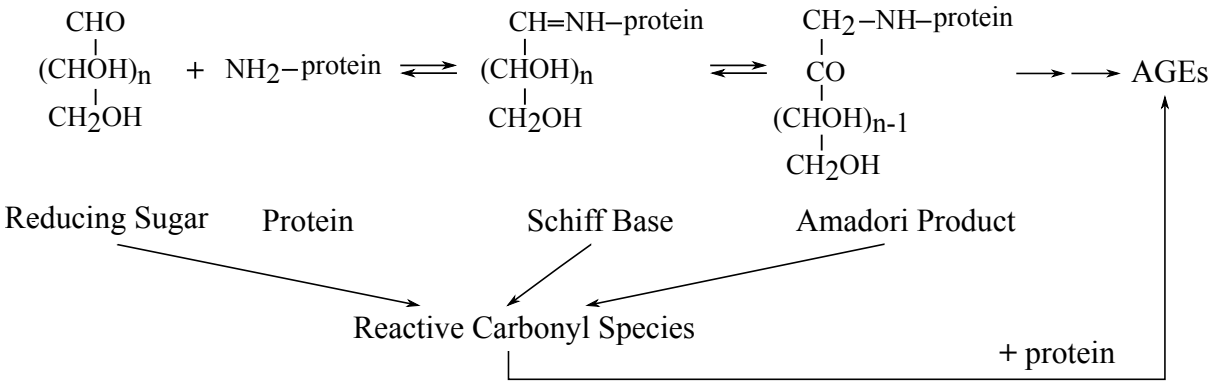


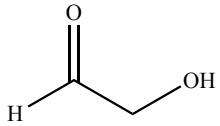
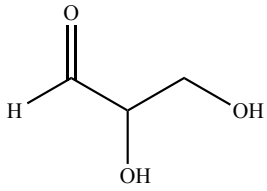
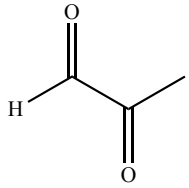
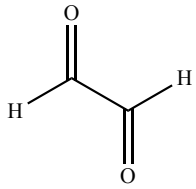
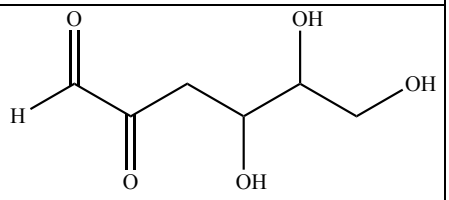
Figure 3-1 A simplified scheme of the glycation reaction, starting from reducing sugar and a protein, via Schiff base and Amadori product, and a cascade of reactions, ultimately forming advanced glycation endproducts (AGEs).

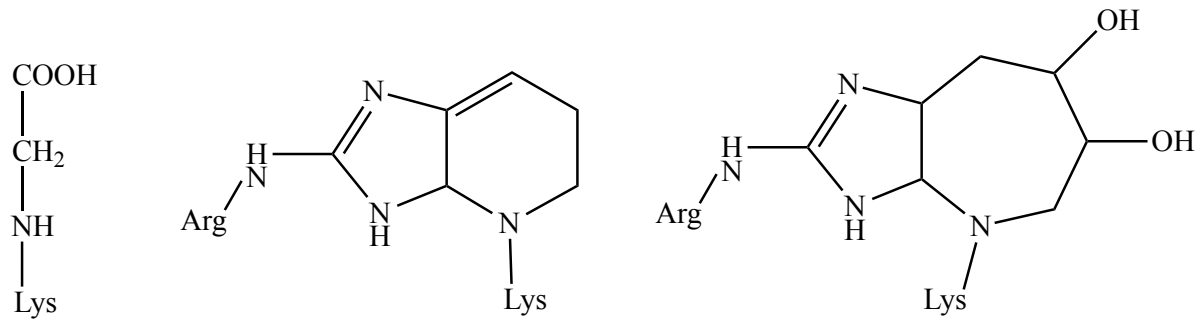
Protein glycation with reducing sugars primarily occurs at ε-amino groups on K and Hyl and to some extent, the guanidino group on R. The presence of amino groups of protein terminal residues can also initiate the glycation reaction with reducing sugars. Glycation products have also been identified on residues other than K, Hyl and R, such as amino groups on H (from reaction with malondialdehyde) [65] and W (from reaction with glucose, likely from reaction with dicarbonyl compounds) [66] and sulfhydryl group on C (from reactions with methylglyoxal, glyoxal and glyceraldehyde) [67][68]. Even though the sulphide on M can be susceptible to nucleophilic attack, the reaction requires stronger electrophiles than reducing sugars and the common carbonyl species generated in the glycation reaction process; therefore, it is not likely to happen.

AGEs generally fall into two categories: monovalent amino acid sidechain modifications, such as N^{ε}-(carboxymethyl)lysine (CML), and divalent crosslinks between two residues, such as pentosidine and glucosepane (Figure 3-2), which are both crosslinks between K and R residues.

Although some AGEs have been identified in native tissues and in model systems (reviewed in section 3.4), and model AGE compounds have been synthesised, there is little known about the reaction steps and glycation intermediates between Amadori products and most AGEs in the glycation process.

Table 3-1 Examples of possible reactive carbonyl or dicarbonyl intermediates generated in the glycation process [69][70].

Reactive intermediates	**Formula**	**Structures**
Carbonyl Species		
Glycolaldehyde	CH_2OH-CHO	
Glyceraldehyde	CH_2OH-CHOH-CHO	
Dicarbonyl Species		
Methylglyoxal (MGO)	CH_3-C(O)-CHO	
Glyoxal (GO)	OHC-CHO	
3-Deoxyglucosone (3-DG)	OHC-C(O)-CH_2-$(CHOH)_2$-CH_2OH	



CML Pentosidine Glucosepane

Figure 3-2 Structures of some well-characterised AGEs: CML, pentosidine and glucosepane.

3.2 Consequences of Glycation

Two biologically relevant consequences can result from glycation. The first one is protein structure modifications and subsequent loss/ alteration of function. For example, decreased cell adhesion and spreading have been observed on glycated collagen matrix [71][72]. The second one is AGEs binding to receptors for AGEs (RAGEs) on cell membrane, triggering abnormal cell signalling pathways and enhancing ROS production, which can further promote the glycation reaction. RAGEs activation and subsequent cellular activities are implicated in contributing to the development of diseases, such as diabetic complications and cancer progression. [50][58]

3.2.1 Consequences of Collagen Glycation

Collagen provides the framework and mechanical strength for many tissues and interacts with many different types of cells. This collagen-cell interaction is essential in maintaining tissue homeostasis. However, collagen glycation has the potential to interrupt the regular collagen-cell interactions and threaten normal tissue function. Additionally, collagen is a long-lived and slow-turnover structural protein, that can accumulate a significant amount of glycation products to potentially induce long term pathological damages.

Consequences of collagen glycation are considered from three aspects here and explained in the following sections. However, those consequences are correlated rather than independent of one another.

Destabilising Collagen Organisation

AGE formation can change the charge distribution on collagen surface. For example, under physiological pH, converting a K residue into CML adds a negatively-charged carboxymethyl moiety on the originally positively-charged sidechain of K (Figure 3-3). This charge distribution alteration has the potential to affect intermolecular electrostatic interactions, critical to collagen fibril assembly and thus, cause misalignments of collagen molecules in fibril assembly.

In addition, glycation by adding a sugar adduct or derivative thereof to a free amino group of collagen K or R changes the bulkiness of their sidechains, which potentially affects the tight packing of collagen molecules and leads to local swelling in collagen fibrils.

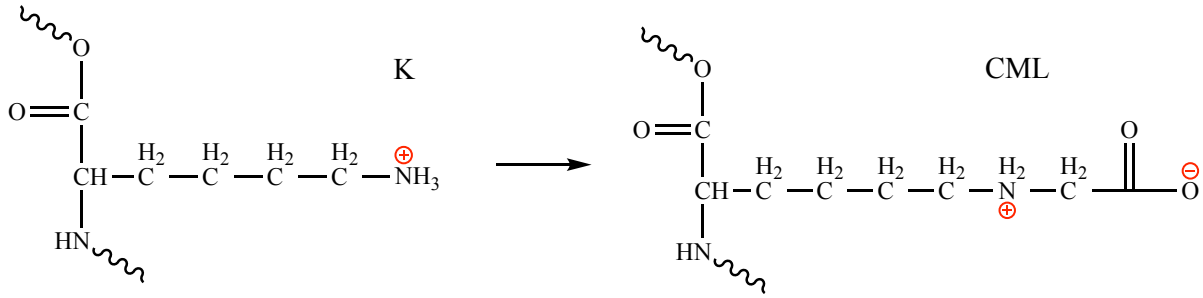


Figure 3-3 Under physiological pH, the K sidechain is positively-charged. When carboxymethyl-lysine (CML) forms, an extra negatively-charged carboxymethyl functionality at the end of the original K sidechain can alter the charge distribution in collagen and the bulkiness of the sidechain.

Collagen glycation has been suggested to target at the same K/Hyl sites in triple helical domain as enzymatic crosslinks and affect the establishment of enzymatic crosslinks [73]. This possible interference of collagen K/Hyl glycation with enzymatic crosslinks may also affect the collagen molecular arrangement in the fibrils and the ability of those arrangements to be restored, e.g. after a tensile force.

Other work (unpublished) in the Duer group has shown that the positions of positively-charged residues K and R are the most conserved of any amino acids in mammalian collagens. Thus, we anticipate that alterations of these residues by glycation can potentially have profound effects not yet discovered.

Impairing Collagen Biological Functions

Glycation products have the potential to hinder binding of ligands at essential biological functional sites on collagen molecules via several possible mechanisms, such as chemical structure modification, steric hinderance and electrostatic repulsion, and thus, impair collagen biological functions.

The sidechain charge alteration caused by glycation can probably influence the collagen-protein interactions in tissue ECMs. It has been reported that KSPGs, with negatively-charged GAG chains, show weaker affinity towards glycated collagen than non-glycated collagen [72]. It is hypothesised that this reduced affinity is contributed by increased repulsion between glycated collagen molecules and the negatively-charged glycan groups on the KSPGs; it is also possible that the binding sites for KSPGs on collagen are modified during the glycation reaction and the loss of some binding sites causes the reduced affinity. As PG-collagen interactions have been proposed to regulate cell-ECM interactions, this weaker affinity between KSPGs and glycated collagen may further affect fundamental cell behaviours, such as adhesion and migration, which is also shown in the study [72].

Decreased cell adhesion and proliferation on glycated collagen versus non-glycated collagen have been observed in other work [71], where it is proposed to relate to the glycation modification of R in specific integrin binding sites on collagen and result in subsequent decrease in integrin-mediated cell adhesion. Nonetheless, the exact mechanisms behind reduced collagen-protein interactions and altered cell behaviours caused by glycation are not clear.

Altering Collagen Fibril Mechanics

AGE crosslinks are believed to underlie changes in mechanical properties of glycated collagen fibres compared to non-glycated collagen, such as increased stiffening and breaking load [74]. However, mechanical properties of glycated collagen are usually measured on a macroscopic level, such as on a piece of glycated tissue or collagen fibres extracted from tissues [75][76], and there is still little known about the effects of collagen glycation and AGE accumulation on mechanical properties at different structural levels of collagen, in particular at the molecular level. In addition, whether the formation of AGE crosslinks is the sole reason underlying

collagen stiffening and other mechanical changes remains to be confirmed. It is important to underline that no study has yet characterised the full distribution of glycation products present in a tissue and so the presence of other glycation products in glycated collagen may yet to be shown to play a role in the changes of collagen mechanics.

3.3 In vivo Relevance of AGEs

The formation and accumulation of AGEs in aged and diseased tissues has long been suggested to contribute to the genesis and development of diseases. However, how exactly the glycation reaction and accumulation of AGEs causes tissue impairment and dysfunction is still poorly understood. This section includes evidence of AGEs in aged and diseased tissues in order to provide an overview of the possible relationship between AGEs and diseases.

Ageing

A number of AGEs have been identified in aged tissues, and the concentrations of AGEs are found to be age-dependent (Table 3-2). For example, carboxymethyl-lysine (CML), one of the most characterised AGEs, has been identified in human lens proteins, skin and tendon collagen [77][78][79]. The CML level in lens protein increases dramatically with age, with CML representing 0.02 to 0.8% of the total lysine residues in lens proteins aged 0 to 79-year-old. This age-dependent accumulation of AGEs has long been proposed to trigger age-related diseases.

However, those studies only focus on one or two AGEs. Whether other AGEs or glycation intermediates are present or accumulate in aged tissues is not clear; neither is the relative concentrations between different AGEs and glycation intermediate products.

Table 3-2 The concentrations of some AGEs in tissues are found to linearly increase with age. CML, carboxymethyl-lysine; CMhL, carboxymethyl-hydroxylysine; pentosidine and glucosepane are glycation crosslinks between Lys and Arg.

AGEs	Tissues	Concentrations	Ref
CML	Human lens protein	~7 mmol CML/ mol Lys at age 80	[77]
	Human skin	~1.5 mmol CML/ mol Lys at age 80	[78]
CMhL	Human skin	~5 mmol CMhL/ mol Hyl at age 80	[78]
Pentosidine	Human dura mater	~250 pmol/ mg collagen around age 100	[80]
Glucosepane	Human skin	~2000 pmol/ mg collagen at age 90	[81]

Diabetes and Diabetic Complications

AGE formation and accumulation occur at a much more accelerated rate under diabetic conditions [30][82][83]. For instance, glucosepane level can go up to 4000 pmol/ mg collagen in diabetic patients versus around 2000 pmol/mg in non-diabetic patients at the same age [84].

The amount of AGEs detected in tissues has been reported to correlate with the severity of diabetic complications by some immunohistochemical studies. For example, the amount of CML and pentosidine in diabetic renal tissues increases with the severity of diabetic nephropathy [85][86]. AGE accumulation potentially causes proteins to be resistant to enzymatic degradation and may account for part of the BM thickening in diabetic nephropathy.

Also, upregulated RAGE expression has been reported in diabetic nephropathy [86] and AGE-RAGE interactions are known to trigger abnormal cellular pathways and lead to cell death [87], suggesting a potential role of AGE-RAGE interactions in diabetic nephropathy and tissue dysfunction.

Many studies suggest that there is a close relationship between AGE accumulation and diabetic complications, but a clear relationship is not yet established. Some hypothesises have been proposed: in diabetes, hyperglycaemia provides excess circulating glucose as glycating agents, so higher glucose concentration can be attributed to the underlying reason for the accelerated glycation rate and elevated AGE levels detected in diabetic patients. Additional and more reactive metabolites can be generated from altered metabolisms in diabetes to facilitate the

glycation reaction, such as fructose, both exogenous and endogenous, and fructose metabolites [88]. For example, elevated polyol pathway, converting glucose to fructose via sorbitol, in diabetes has been suggested to correlate with the development of diabetic neuropathy [89].

However, few studies have mentioned the co-presence of glycation intermediate products and examined their contribution towards diabetic complications and tissue failure. Most research published to date only focuses on one or two AGEs at a time without an assessment of other AGEs and glycation intermediate products. Therefore, whether the detrimental effects associated with diabetic complications is contributed to what extent by AGEs and to what extent by the glycation intermediates is not known.

Cancer

The concentration of AGEs was identified to be elevated in cancerous tissues and correlated with aggressiveness of cancer. Higher CML level is observed in prostate cancer tissues compared to non-cancerous tissues, and it is higher in malignant tumour tissues than benign tumour tissues [59]. Moreover, RAGEs are found overexpressed in a variety of cancers, such as in prostate, breast and lung cancers, suggesting both AGEs and AGEs-RAGEs interactions play a role in cancer genesis and progression. [58][90]

Cancer cells usually uptake more glucose than healthy cells to generate the energy for their rapid proliferation, leading to a higher rate of glycolysis [91]. Alterations of several other metabolic pathways in cancer metabolism are also under investigation, such as increased pentose phosphate pathway activity for antioxidant and nucleotide production to support cancer cell survival and invasion. We propose here that this altered cancer cell metabolism might provide a more diverse collection of glycating agents to the surrounding ECM than present in normal tissues, resulting in more vigorous glycation reactions.

Diabetic patients are found to associate with a higher risk of some cancers, such as liver and pancreas cancers, according to some epidemiological studies [92]. Although the exact underlying mechanism for this association between diabetes and cancer remains unclear, glycation induced changes in diabetes/hyperglycaemia, including elevated AGE accumulation, AGE-RAGE binding and consequential alteration of cellular pathways, ROS overproduction,

etc., all together are speculated to generate a pro-cancerogenic condition and link diabetes and cancer. [57][93][94][95]

3.4 Glycation Studies in Literature

Glycation has been intensively investigated for decades in different systems to understand the chemistry behind glycation, aiming to (1) use glycation products as biomarkers for disease and (2) find targeted therapeutic methods to prevent the effects of glycation and eventually reverse them. However, due to the complex chemistry of glycation, glycation chemistry is still not yet fully understood, and AGEs generated from this complicated reaction pathway are far from fully identified. Additionally, no one has studied the glycation intermediate products.

In order to study and identify unknown glycation products, especially glycation intermediate products, a list of glycation products identified in both in vivo systems and physiological in vitro systems over the past decades is reviewed in this section to compare glycation products identified in this work and previously in literature. Their names are numbered, and their structures are included in the Appendix. It is not surprising that most of the glycation products identified to date are AGEs; little is known between the transition from Amadori products to AGEs. Later in this section, commonly used analytical techniques for glycation studies are evaluated to identify their advantages and disadvantages and whether they are suitable for studying glycation intermediate products.

3.4.1 Glycation Products Identified in In vivo Studies

Glycated proteins, haemoglobins, were first detected in diabetic patients' blood in 1969 [96]; but not until the early 1980s, glycation of long-lived proteins was brought to attention and proposed to contribute to tissue ageing and the development of diabetic complications. Since then, a number of glycation products have been identified in in vivo collagen-rich tissues, mostly aged or diabetic tissues.

AGEs that have been detected in in vivo tissues are summarised in Table 3-3, including the techniques utilised to identify AGEs. Usually, biological samples are digested, either acid hydrolysis or enzymatic digestion, before using HPLC to separate and purify the species of interest, and then, the species of interest is confirmed by HPLC coupled fluorescence

spectroscopy or mass spectrometry or both. Sometimes, the HPLC separated AGEs are further studied by NMR spectroscopy to reveal chemical structures. Some research has been performed on biological fluids, such as urine and blood, probably because they are easier to sample and analyse, and those AGEs may also present in collagenous tissues.

Table 3-3 Glycation products have been identified in different in vivo systems using different analytical techniques. Many of them are studied in in vitro systems to obtain pure products and characterise the pure products prior to moving on to in vivo systems. Most of them are detected to be at a relatively low level in healthy or young tissues and higher quantities in aged or diabetic tissues. Their chemical structures and ^{13}C NMR chemical shifts (from solution-state NMR) assigned in the literature are included in Appendix. AA, amino acid involved forming the glycation product; SIM-GC/MS, selected ion monitor mode of gas chromatography and mass spectroscopy; ICS, immunohistochemical staining.

No.	Glycation Products	AA	In vivo Systems	Detection Methods
AGEs - Amino Acid Sidechain Modifications				
1	N^{ε}-(carboxymethyl)lysine (CML)	Lys	Normal and aged human lens proteins [77] Human skin collagen of different ages [78]	SIM-GC/MS SIM-GC/MS
2	N^{ε}-(carboxymethyl)hydroxylysine (CMhL)	Lys	Human skin collagen of different ages [78]	SIM-GC/MS
3	N^{ε}-(carboxyethyl)lysine (CEL)	Lys	Various normal and diabetic human and animal tissues [97] Healthy human and diabetic rat blood [98]	 LC-MS/MS
4	Pyrraline	Lys	Alzheimer's disease tissue sections [53] Induced diabetic rats [99]	ICS
5	GA-pyridine	Lys	Human aorta [100]	ICS
6	Methylglyoxal hydroimidazolone (MG-H1)	Arg	Various normal and diabetic human and animal tissues [97] Healthy human and diabetic rat blood [98]	 LC-MS/MS
7	Glyoxal hydroimidazolone (G-H1)	Arg	Various normal and diabetic human and animal tissues [97] Healthy human and diabetic rat blood [98]	 LC-MS/MS
8	3-Deoxyglucosone hydroimidazolone (3DG-H)	Arg	Healthy human and diabetic rat blood [98]	LC-MS/MS
9	Argpyrimidine	Arg	Diabetic tissues [101] Healthy human and diabetic rat blood [98]	ICS LC-MS/MS
10	Tetrahydropyrimidine (THP)	Arg	Diabetic tissues [102]	ICS
11	S-(carboxymethyl)cysteine (CMC)	Cys	Diabetic patients blood and urine [103] Rat liver mitochondrial [104]	LC-MS GC-MS

12	S-(carboxyethyl)cysteine (CEC)	Cys	Diabetic patients blood and urine [103]	LC-MS
AGEs - Crosslinks				
13	Pentosidine	Lys, Arg	Human skin, aged and diabetic [49] Human skin, diabetic and non-diabetic [105]	HPLC-FLUO
14	Glucosepane	Lys, Arg	Various normal and diabetic human and animal tissues [97]	
15	Methylglyoxal-lysine dimer (MOLD)	Lys	Aged lens protein and skin collagen [106] Healthy human blood [98]	LC/MS
16	Glyoxal-lysine dimer (GOLD)	Lys	Aged lens protein and skin collagen [106]	LC/MS
17	Lysine-dihydropyridinium (K2P)	Lys	Human lens protein [107]	LC-MS/MS
Intermediates				
18	Fructose-lysine (FL)	Lys	Normal and aged human lens proteins [77] Human skin collagen of different ages [78] Various normal and diabetic human and animal tissues [97]	SIM-GC/MS SIM-GC/MS

3.4.2 Glycation Products Identified in In vitro Physiological Model Systems

Because of the complexity in studying glycation in in vivo systems, much research has done in in vitro model systems with a wide selection of glycation substrates, including protected amino acids, small protein and pure collagen, and a variety of glycating agents, such as hexose, pentose, oxoaldehydes. A list of glycation products identified in different in vitro systems is summarised in Table 3-4.

Table 3-4 shows that most glycation products are identified in model systems using either an amino acid mixture, such as N-acetyl-lysine and N-acetyl-arginine, or small proteins with known sequences or functions, such as bovine serum albumin (BSA). These relatively small and soluble glycation substrates provide access to glycation products in relatively short reaction time and the glycation products can be characterised easily by a number of analytical techniques. However, to some extent, these in vitro systems lack the biological relevance of what happens in diseased tissues. Formation of glycation products in collagen fibrils, the focus of this work, can be prohibited by a number of structural factors/restrictions posed by the complex collagen hierarchical structure, which is absent in amino acid solutions or small plasma proteins like human serum albumin (HSA). For example, the formation of glycation crosslinks requires two reactive residues that are spatially close and sterically able to interact, either within the same collagen molecule or between collagen molecules, which are less likely in collagen fibrils than small and flexible proteins. Therefore, it is not unexpected that some of the glycation products identified in vitro are not identified in glycated collagen fibrils.

Table 3-4 Glycation products have been identified in different physiological in vitro model systems using different analytical techniques. Their chemical structures and ^{13}C NMR chemical shifts (from solution-state NMR) assigned in the literature are included in Appendix. AA, amino acid involved forming the glycation product; Glc, glucose; BSA, bovine serum albumin; Rib, ribose; GO, glyoxal; MGO, methylglyoxal; 3-DG, 3-deoxyglucosone; 3-DP, 3-deoxypentosulose; HSA, human serum albumin.

No.	Glycation products	AA	In vitro Model systems	Detection Methods
AGEs - Amino Acid Sidechain Modifications				
19	N-(carboxymethyl)arginine (CMA)	Arg	Collagen and Glc [108]	MS, ^{1}H NMR
AGEs - Crosslinks				
20	Pentosinane (precursor of pentosidine)	Lys, Arg	BSA and Glc/Rib/xylose/arabinose [109]	NMR
21	Glyoxal-derived-imidazoline-crosslink (GODIC)	Lys, Arg	BSA and GO/Glc [110]	HPLC-UV, NMR
22	Methylglyoxal-derived-imidazoline-crosslink (MODIC)	Lys, Arg	BSA and MGO/Glc [110]	NMR
23	DOGDIC	Lys, Arg	BSA and 3-DG/3-DP/Glc/Rib/xylose/arabinose [109]	NMR
24	DOPDIC	Lys, Arg	BSA and 3-DG/3-DP/Rib/xylose/arabinose [109]	NMR
25	Arg-hydroxy-triosidine	Lys, Arg	Human corneas and GA [111]	HPLC-UV-MS NMR
26	Vesperlysine A, B and C	Lys	BSA and Glc [112]	NMR
27	3-Deoxyglucosone-lysine dimer (DOLD)	Lys	Hippuryl-lysine and 3-DG [113]	HPLC NMR
28	GALA	Lys	N^{α}-t-Boc-Lys and Rib/Glc [114]	NMR
29	GOLA	Lys	N^{α}-t-Boc-Lys and Rib/Glc [114]	NMR
30	Glucose lysine dimer (GLUCOLD)	Lys	BSA and Glc [115]	NMR
31	Lys-hydroxy-triosidine	Lys	Human corneas and GA [111]	HPLC-UV-MS NMR

3.4.3 Analytical Techniques

Another problem raised by the majority of existing studies is that most of them, to some extent, are targeted to identify specific glycation products rather than mapping all possible glycation products in a model system, due to the inherent limitations of the analytical techniques used in those studies. This section evaluates the advantages and disadvantages of the main analytical techniques used to date to measure glycation products of collagen. [116][117]

Liquid Chromatography-Mass Spectroscopy (LC-MS)

AGEs are mostly identified by a combination of chromatography and mass spectroscopy, traditionally LC-MS and more advanced HPLC-MS/MS. These techniques require the least amount of sample (around 10μg protein) compared to other techniques discussed below and can produce quantitative results. However, these LC-MS-based techniques require standard compounds of each glycation product of interests to be made and examined before testing a complex biological mixture. However, concerns exist regarding preanalytical sample processing associated with this technique. Even though enzyme digestion is replacing traditional acid hydrolysis of collagenous tissues, revealing some acid-labile glycation products previously unidentified, there is no verified protocol to prepare protein samples to map all glycation products. Insufficient release of glycation products can be a problem when glycation products hinder proteolysis sites. Moreover, errors can be introduced during ionisation in the MS step, because the ionisation of some glycation products can cause oxidation and/or dehydration. For example, fructose-lysine (FL) can degrade to CML during ionisation [118].

Immunohistochemical staining by antibodies

Immunohistochemical staining has been used to map spatial distribution of some specific glycation products in native tissues [101].

This technique has a few drawbacks. First, the antibodies used are often not standardised, and different laboratories tend to raise their antibodies. Second, the specificity of antibodies against the targeted glycation product is questionable. Antibodies can cross-react with substances that share similar molecular structures with their targets. For example, an antibody against CML, 6D12, is found to react with CEL as well. [118][119] Third, the detection also depends on the

penetration depth of the antibody into the tissue sample, which can result in unreproducible results. Moreover, quantification methods relying on antibodies, such as ELISA assays, often do not offer an absolute concentration of substrate but only relative comparisons between samples, which makes it difficult to compare its results with results from other techniques.

Fluorescence

Some AGEs are fluorescent, offering the opportunity to monitor their presence by relatively simple and non-invasive fluorescence detection.

However, most AGEs are not fluorescent and therefore cannot be detected. Moreover, fluorescence emission from different fluorescent AGEs overlap with each other extensively so, fluorescence spectra are not able to identify nor quantify specific glycation products, but only able to follow some part of the glycation process qualitatively. Other fluorophores, such as oxidation products like dityrosine, may coexist in biological samples, further interfering with glycation fluorescence measurement.

Nuclear Magnetic Resonance (NMR) Spectroscopy

Some AGEs are identified by solution-state NMR spectroscopy. Chemical structures of AGEs are analysed based on their ^{1}H or/and ^{13}C chemical shifts. Most of the AGEs generated in in vitro incubation are characterised by solution-state NMR spectroscopy; some AGEs in vivo tissues are purified by HPLC and then identified by solution-state NMR spectroscopy.

In addition, solid-state NMR spectroscopy is considered as an emerging technique in this field and has the advantage of simplicity and minimal sample preparation – lyophilisation.

3.5 Summary: A Novel In vitro Glycation Model System

The glycation reaction and the accumulation of AGEs in tissues are widely accepted to associate with ageing, the development of diabetic complications and cancer genesis. A huge amount of work has been done aiming to unveil the mechanism behind AGE accumulation and disease pathologies, but the relationship between the accumulation of AGEs and disease genesis and development is still not fully understood.

Most research to date has focused on measuring only one or two AGEs under diseased conditions with special attentions on glycation crosslinks. However, all the deleterious effects caused by glycation cannot possibly be attributed to one or two AGEs. Considering the complex chemistry of the glycation process, a variety of glycation products can be generated at different stages of the glycation process, intermediates and endproducts, and are likely to all be present at some concentration in aged and diseased tissues, and therefore, consequences of glycation are probably caused by a collective effect of all AGEs and glycation intermediates present at glycation equilibria.

However, little is known about the glycation intermediate products in biological systems, not to mention the relative amount of different AGEs and different glycation intermediates. One or two AGEs present in minor quantities are not likely to cause all the detrimental effects described in the previous sections. More studies are needed to identify glycation intermediate products, study the effects of glycation intermediate products and assess the relative significance between AGEs and the glycation intermediates.

In order to mimic the glycation reaction and identify biologically relevant glycation products in vitro, the glycation model system needs to be optimised to best represent in vivo situation. Therefore, a novel model system is used in this work in order to produce glycation products that are more representative of in vivo situations, and solid-state NMR spectroscopy is used to profile all glycation products present at the glycation equilibria, with a focus on glycation intermediate products generated in this model system.

Choice of the Glycation Substrate

Collagen dominated in vitro ECMs are used as the primary glycation substrate in this project. It has been shown that in vitro ECM from osteoblasts mimic in vivo bone tissues [120]. The in vitro ECM bridges the gap in understanding of glycation chemistry between amino acid mixtures and small proteins and native tissues.

Choice of the Glycating Agent

In theory, any chemical with a carbonyl functional group, ketones or aldehydes, can potentially react as a glycating agent. However, two factors need to be taken into consideration. The first

one is the biological relevance of the carbonyl species, i.e. whether they are generated by metabolism and likely to be present in ECMs in vivo. The second one is their glycation rates. Glycating agents with rapid glycation rate are more likely to generate detrimental effects in a short time scale. It is also desirable to use an in vitro model glycating agent where the product equilibrium distribution is reached within a manageable timescale.

Glucose is clearly the most biologically relevant glycating agent; however, glucose glycates very slowly and takes long time to reach the equilibrium distribution of glycation products, and is therefore, not amenable to study in this project. Ribose-5-phosphate (R5P), in the end, is chosen as the primary glycating agent in this project. R5P glycates faster than glucose, fructose and ribose [121] so that it can give access to the glycation equilibrium within reasonable timescales. It is a less studied but versatile metabolite involved in multiple metabolic pathways, and its production is suggested to be upregulated in cancerous tissues to support rapid proliferation of cancer cells [91], and so we propose that R5P will be an important glycating agents where there is cell necrosis in cancer.

Choice of the Major Analytical Technique

SSNMR spectroscopy is chosen as the primary tool to assess collagen glycation product distribution for several reasons. First, it requires minimal sample preparation – only lyophilisation and best preserves proteins in their native states. This feature minimises the risks of introducing artefacts in the results. Second, SSNMR allows observation all glycation products generated in a single sample and therefore clear assessment of major glycation products in each case.

4 Solid-State Nuclear Magnetic Resonance (SSNMR) Spectroscopy

Solution-state NMR spectroscopy has become a routine characterisation technique for small molecules and soluble proteins, while SSNMR spectroscopy is far less exploited. SSNMR and latest developments in dynamic nuclear polarisation (DNP)-coupled SSNMR pose significant advantages for studying insoluble macromolecules.

In this project, SSNMR is the major analytical tool, and a large component of this project is analysing SSNMR spectra to identify glycation products and explore atomic-level structural changes of insoluble collagenous materials.

Essential concepts of SSNMR are briefly reviewed in this chapter, followed by principles of routine SSNMR experiments, focusing on SSNMR spectra interpretation. Complete understanding of NMR spectroscopy can be found in textbooks, *Understanding NMR Spectroscopy* by James Keeler [122], *Introduction to Solid-State NMR spectroscopy* by Melinda Duer [123] and *Solid-State NMR: basic principles & practice* by David Apperly, Robin Harris and Paul Hodgkinson [124], etc.

4.1 The Vector Model

When a sample is placed in a magnetic field, B_0, a bulk magnetisation, $\boldsymbol{M}$, will be generated under the effect of B_0, as shown in Figure 4-1. $\boldsymbol{M}$ is a vector sum of all individual magnetic moments, μ_i, arising from each nucleus in that sample.

$$\boldsymbol{M} = \sum_i \boldsymbol{\mu_i} \qquad \text{(Equation 4-1)}$$

The individual magnetic moment is related to the nuclear spin, I_i, of the *i*th nucleus and the gyromagnetic ratio of the nuclear, γ.

$$\boldsymbol{\mu_i} = \gamma \boldsymbol{I_i} \qquad \text{(Equation 4-2)}$$

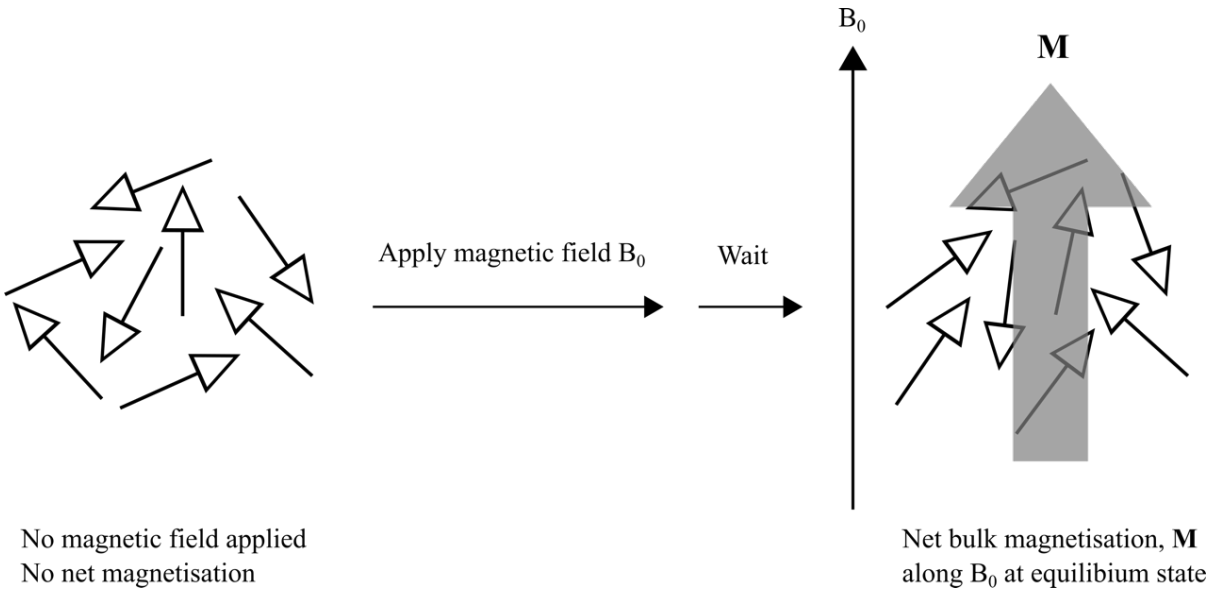


Figure 4-1 Spins in a sample are randomly orientated without an external magnetic field, and become more organised and generate a bulk magnetisation, $\boldsymbol{M}$, under the effect of the applied magnetic field, B_0. $\boldsymbol{M}$ is parallel to B_0.

Nuclear spins precess about the static magnetic field, B_0, at a constant rate, called Larmor frequency, ω_0. This is determined by the magnitude of B_0 and the gyromagnetic ratio, γ, of the nuclear.

$$\omega_0 = -\gamma B_0 \qquad \text{(Equation 4-3)}$$

In pulsed NMR experiments, radiofrequency (rf) pulses, oscillating along the direction perpendicular to B_0, flips $\boldsymbol{M}$ away from equilibrium. Considering a simple case when the frequency of the rf pulse is the same as the Larmor frequency of the nuclear ($\omega_{rf} = \omega_0$), $\boldsymbol{M}$ will flip an angle of θ_{rf} during time τ_{rf}.

$$\theta_{rf} = \omega_{rf}\tau_{rf} = -\gamma B_{rf}\tau_{rf} \qquad \text{(Equation 4-4)}$$

where B_{rf} is the rf field and ω_{rf} is the frequency of B_{rf}.

Conventionally, a 90° pulse in NMR experiments means turning the magnetisation 90° away from the vertical axis, where B_0 lies, to the transverse plane (Figure 4-2). The magnitude of the transverse magnetisation during its recovery to equilibrium is what detected over time in pulsed NMR experiments.

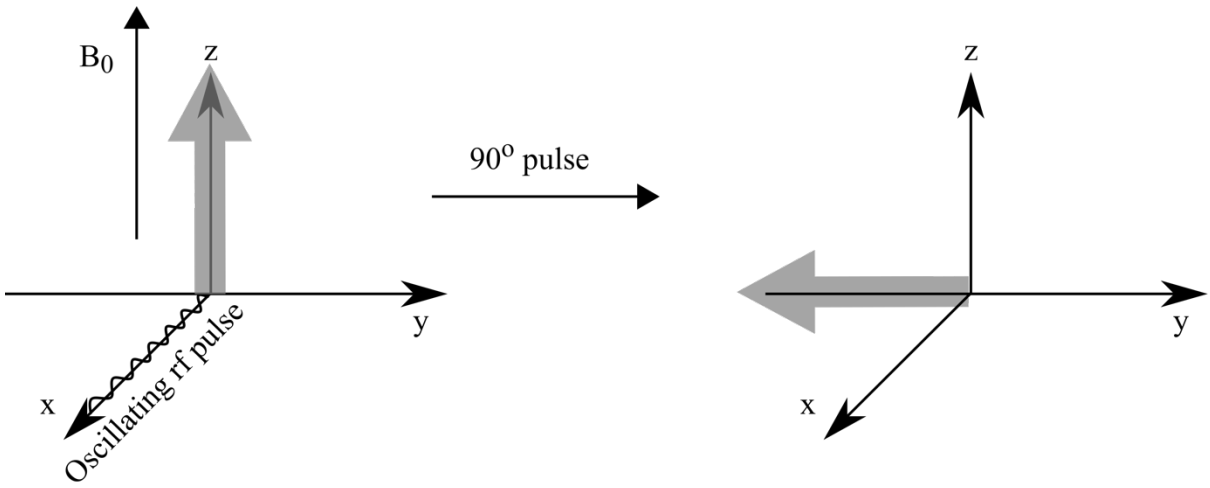


Figure 4-2 A 90° pulse generated by a rf pulse oscillating along the x axis and perpendicular to B_0, along the z axis, flips the bulk magnetisation, represented by the filled grey arrow, from equilibrium to transverse plane, along the y axis.

4.2 Types of Interactions

In a spin system, nuclear spins interact not only with the static magnetic field, but also with each other. For a spin system containing interacting spins, the overall interactions include but are not limited to the Zeeman interaction, the shielding interaction and the dipolar interactions, including both homonuclear and heteronuclear dipolar interactions. All three interactions will be discussed separately in detail in the following sections as all are relevant to the experiments presented in this thesis.

For spins with $I > \frac{1}{2}$, there are also quadrupolar interactions, as such nuclei possess a non-zero electric quadrupole moment which interacts with surrounding electric field gradients. However, nuclei studied in this project are spin half nuclei, $I = \frac{1}{2}$, and thus, quadrupolar interactions will not be discussed in this thesis. Scalar coupling (or J coupling) is a further nuclear spin interaction in which pairs of nuclear spins are coupled via through-bond interactions, it is not exploited in the solid-state NMR experiments used in this project; therefore, it will not be discussed in detail either.

4.2.1 Zeeman Interaction

The Zeeman interaction is the interaction between the spin system and the applied field, which is the dominant interaction in a spin system. It is determined by the magnitude of B_0 and gyromagnetic ratio, γ, of the nuclei, meaning it is the same for all spins of the same nuclei in a given field.

4.2.2 Shielding Interaction

The shielding interaction describes the interaction between the static magnetic field, B_0, and the local magnetic field generated by surrounding electrons of a nucleus, ΔB, which is proportional to B_0. The nucleus is shielded from B_0 by ΔB and therefore, experiences a net magnetic field, $B_0 - \Delta B$. As the electron cloud around nuclei is general not spherically symmetric, ΔB is different depending on the relative orientation of the electron distribution with respect to B_0, resulting in so-called shielding anisotropy. The anisotropy of the shielding interaction is described mathematically by the shielding tensor, $\boldsymbol{\sigma}$, a second rank tensor whose three principal values are designated σ_{11}, σ_{22} and σ_{33}.

In modern NMR spectrometers, chemical shifts are introduced, converting from the absolute frequencies (in Hz) of NMR signals to chemical shift (in ppm, parts per million) to remove the field strength dependence of the shielding interaction:

$$\delta(ppm) = \frac{actual\ fequency\ (Hz) - standard\ reference\ frequency\ (Hz)}{standard\ reference\ frequency\ (Hz)} \times 10^6 \qquad \text{(Equation 4-5)}$$

Within a solid sample, molecules, and thus their associated electron distributions, are in all possible orientations, so that the shielding anisotropy results in a different chemical shift for each molecular orientation. All chemical shifts arising from different molecular orientations combine together as a "powder pattern" spectrum (Figure 4-3), which the width of the power pattern being related to the chemical shift anisotropy (CSA). CSA is used to describe the orientation dependence of chemical shifts.

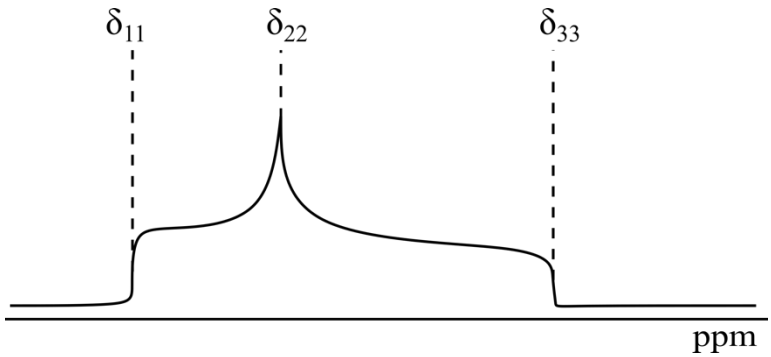


Figure 4-3 A typical powder pattern spectrum showing three principal values of the shielding tensor where $\sigma_{11} \geq \sigma_{22} \geq \sigma_{33}$. The actual shape of the powder pattern depends on the symmetry of the nuclei.

4.2.3 Dipolar Interaction

As mentioned previously, each nucleus in a molecule can generate a local magnetic field. Apart from interacting with B_0, these local fields also interact with fields generated by nearby nuclei. This interatomic interaction is called the dipolar or dipole-dipole interaction, which occurs through space and is independent of bonds between dipoles.

Similar to the shielding interaction, the strength of the dipolar interaction also depends on orientation, more specifically, the angle between the internuclear vector and B_0, θ. It also depends on gyromagnetic ratios of both spins involved, γ_I and γ_S, and internuclear distance, r_{IS} (Figure 4-4). The dipolar interaction, d_{IS}, between spin I and spin S can be calculated by

$$d_{IS} = \frac{\hbar\mu_0}{4\pi}\frac{1}{r_{IS}^3}\gamma_I\gamma_S\frac{3\,cos^2\,\theta-1}{2} \qquad \text{(Equation 4-6)}$$

$\hbar$ is reduced Planck constant, μ_0 is the permeability of free space.

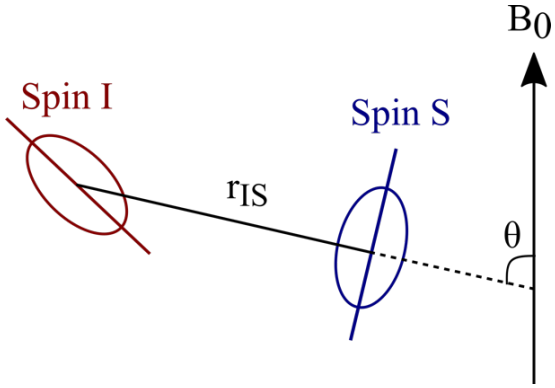


Figure 4-4 The dipolar interaction between spin I and spin S depends on their internuclear distance, r_{IS} and their orientation relative to B_0, θ.

The dipolar interaction is very useful in determining interatomic distances. However, the dipolar interaction also causes extensive signal broadening and potential signal overlapping, especially in biological samples where there are too many ^{1}H-^{1}H and ^{1}H-^{13}C interactions with various internuclear distances and orientations. Therefore, decoupling techniques (will be described in Section 4.3.2) are often used in SSNMR experiments to decouple undesired dipolar interactions and selectively observe informative dipolar interactions.

4.3 Techniques in SSNMR

4.3.1 Magic Angle Spinning (MAS)

In solution-state NMR, above-mentioned anisotropic interactions are averaged, due to rapid isotropic molecular motion in solution, to their isotropic value resulting in sharp signals in solution-state NMR spectra. However, in the solid-state, molecular motion is much slower, and typically, not isotropic, diminishing this averaging effect observed in solution-state NMR. Therefore, SSNMR spectra generally contain broad and overlapping signals from the effects of anisotropic interactions.

A useful averaging effect can be achieved in SSNMR by mechanically spinning samples at an angle of $54.74°$ relative to the applied magnetic field, called the magic angle, which is equivalent to the body diagonal of a cube (Figure 4-5). At the magic angle, the $3\cos^2\theta - 1$ orientation dependence of nuclear spin interactions is averaged to zero. Therefore, spinning at the magic angle can remove the effects of anisotropic interactions. In practice, the spinning rate should be at least three times bigger than the magnitude of the broadening interaction to fully remove its effects from the spectrum. Otherwise, spinning sidebands, a set of signals on both sides of a central signal with multiples of the MAS rate apart, appear on SSNMR spectra with MAS (Figure 4-6).

Rotational resonance occurs when $\omega_d^{iso} = n\omega_r$, where ω_d^{iso} is the difference in isotropic frequencies of two signals, ω_r is the MAS rate and n is an integer. Under the effects of rotational resonance, the dipolar interaction averaged out by MAS is reintroduced, causing signal broadening and splitting. In two-dimensional SSNMR experiments, it is used to reintroduce the dipolar interactions in order to detect homonuclear dipolar coupled spin pairs.

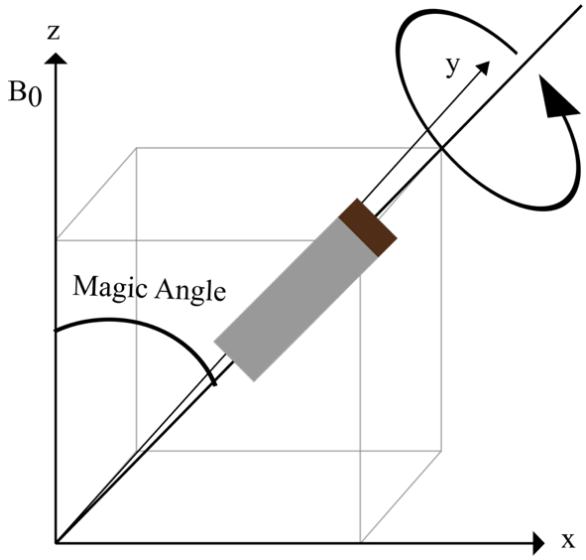


Figure 4-5 Samples are packed in rotors and spun at an angle of 54.74°, called the magic angle, relative to the direction of B_0, along z-axis to remove anisotropic interactions. This angle is at the body diagonal of a cube.

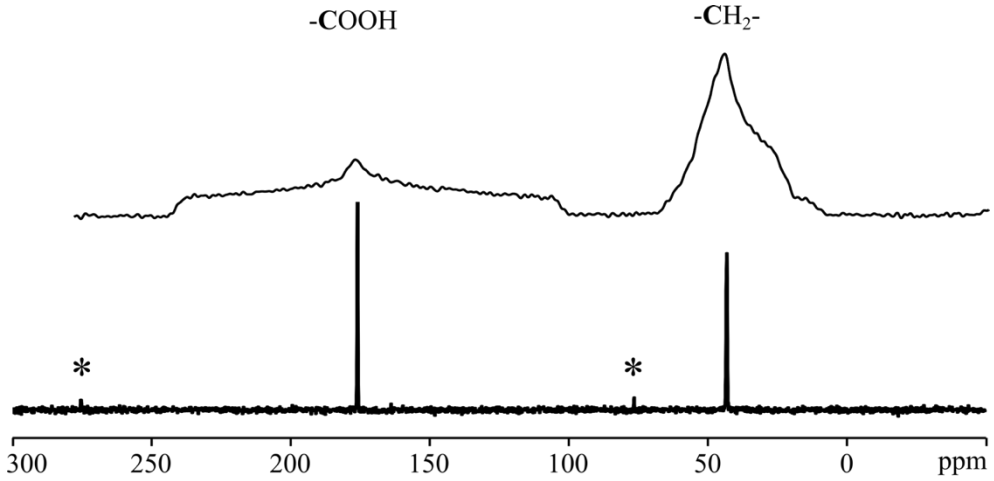


Figure 4-6 ^{13}C SSNMR spectra of glycine. Top: static spectrum. Bottom: MAS at 10 kHz, with asterisks indicating spinning sidebands with 10 kHz apart from the central signal.

Choice of MAS Rate

In this project, a MAS rate at 10 kHz is used on most samples, unless otherwise stated. A MAS rate is carefully chosen to avoid overlapping between spinning sidebands and signals of interest and rotational resonance effects. Moreover, it has to be reasonably fast to remove some effects of anisotropic interactions but is limited by the diameter of the probe used, 4 mm, to a maximum of 15 kHz. Although this MAS rate is not sufficient to remove the effect of ^{1}H-^{1}H homonuclear dipolar interactions or dipolar couplings of ^{13}C spins with large pools of ^{1}H spins,

because of ^{1}H's large gyromagnetic ratio, good quality ^{13}C SSNMR spectra can be acquired from ^{1}H-rich biological samples.

4.3.2 Decoupling

Biological samples typically suffer large ^{1}H-^{1}H homonuclear dipolar coupling and ^{1}H-X, eg. ^{1}H-^{13}C, heteronuclear dipolar couplings, causing significant signal broadening in ^{1}H and ^{13}C spectra respectively. To reduce the effects of these interactions and improve spectral resolution, decoupling techniques are commonly used in SSNMR. Due to ^{1}H's large gyromagnetic ratio, it can be challenging to remove ^{1}H-^{1}H homonuclear dipolar coupling in ^{1}H spectra; however, decoupling ^{1}H-X heteronuclear dipolar interactions is relatively straightforward to achieve in SSNMR.

Briefly, decoupling ^{1}H from other less abundant nuclei requires high power rf pulses applied close to ^{1}H resonance frequency. The effect of those rf pulses is continuously flipping proton magnetisation, resulting in the time-averaged heteronuclear dipolar coupling to be zero.

There are a number of decoupling sequences widely used in SSNMR, e.g. continuous wave (CW), two pulse phase modulation (TPPM) [125] and small phase incremental alteration (SPINAL64) [126]. The latter two schemes are extensively used in experiments described in this thesis.

4.3.3 Cross Polarisation (CP)

CP experiments are more commonly used than direct polarisation (DP) experiments in SSNMR where the observed spin is a low abundance nucleus such as ^{13}C, and where there are abundant spins with large gyromagnetic ratios, such as ^{1}H. In CP experiments, magnetisation from the abundant spin, mostly ^{1}H, is transferred to a less abundant or dilute spin X, such as ^{13}C or ^{15}N, via their mutual heteronuclear dipolar coupling, resulting in enhanced signal intensity for the low abundance nucleus. A theoretical signal intensity enhancement is a factor of γ_H/γ_X; for example, in ^{1}H-^{13}C CP, the theoretical ^{13}C signal intensity enhancement is $\gamma_H/\gamma_C \approx 4$ times. However, this cannot be achieved in actual SSNMR experiments due to relaxation of both the spin-locked proton transverse magnetization and the developing ^{13}C transverse magnetization during the contact pulses (more specifically, $T_{1\rho}$; discussed in Section 4.4).

Dilute nuclei tend to have long T_1 relaxation time (explained in Section 4.4), meaning the waiting time between consecutive scans, called the recycle delay, in DP experiments is long. However, in CP experiments, recycle delay depends on the ^1H T_1 rather than X T_1, which is typically much shorter. This shortens the required recycle delay and allows more acquisitions per unit time. This plus the increased signal intensity for each scan results in shorter experiment times to reach a desired S/N ratio.

Pulse sequence of CPMAS is shown in Figure 4-7. The abundant spin, I, is firstly excited using a 90° pulse. Then a pair of cross polarisation pulses, also called contact pulses, is applied on both spins for a length of time, known as the contact time. Until an optimal contact time, signal intensity of S builds up with contact time, and afterwards signal intensity starts to decline due to $T_{1\rho}$ relaxation. Finally, signals are acquired on the dilute spin, S, with decoupling pulses on I.

Effective magnetisation transfer between two spins, spin I and spin S, only happens when they are dipolar-coupled, and the Hartmann-Hahn condition for the nutation frequency of the contact pulses is matched: [127]

$$\omega_I = \omega_S \qquad \text{(Equation 4-7)}$$

$$\gamma_I B_I = \gamma_S B_S \qquad \text{(Equation 4-8)}$$

CPMAS experiments, i.e. CP under MAS conditions, are routinely carried out on all samples in this project to show basic molecular structural information on collagenous samples, such as amino acid composition and protein secondary structure. However, information provided by CPMAS experiments is not quantitative in that the CP signal intensity depends on not only the number of nuclei but also the efficiency of magnetisation transfer during CP and the rate of $T_{1\rho}$ relaxation. For example, in ^1H-^{13}C CP, ^{13}C in close spatial proximity to low-mobility ^1H benefits from high CP efficiency and gives intense signals, while ^{13}C in high-mobility groups, such as methyl groups with rapidly rotating ^1H, can suffer low CP efficiency, because the methyl group motion partially averages the methyl ^{13}C-^1H dipolar coupling and results in weak signals. So, it needs longer contact time to build up signal intensities. More importantly, CP works as an essential step for many 2D SSNMR MAS experiments to generate the initiate e.g. ^{13}C transverse magnetisation in the experiment (will be described in Section 4.5).

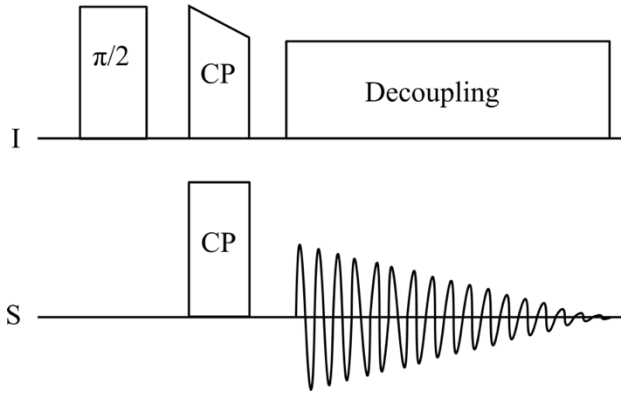


Figure 4-7 A typical CPMAS sequence. Transverse magnetisation is generated by a 90° pulse on the abundant spin I. Then, contact pulses are applied on both spins to transfer magnetisation from spin I to dilute spin S. A ramp pulse is usually used to broaden Hartmann-Hahn condition and gain signal intensity [128]. Finally, S spectrum is acquired with decoupling on I.

4.4 Relaxation

After transverse magnetization is generated, whether by DP or CP, the spontaneous process of restoring the equilibrium along B_0 is called relaxation. Relaxation is caused by fluctuations of local fields experienced by nuclei, under the influence of molecular motion and interactions described in Section 4.2. The restoration of the z-component of $\boldsymbol{M}$ is through longitudinal relaxation, also known as spin-lattice relaxation, with time constant T_1. T_1 determines the necessary recycle delay in NMR experiments. An additional relaxation time becomes relevant in pulsed SSNMR experiments, which is $T_{1\rho}$. This describes the relaxation under the influence of the RF pulses.

The restoration of the equilibrium transverse component of $\boldsymbol{M}$, i.e. 0, is through transverse relaxation, also known as spin-lattice relaxation, with time constant T_2. T_2 is related to the broadness of signals. Slower T_2 is, narrower the signal is. Therefore, relaxation can be seen as an indicator of molecular motion.

4.5 Dynamic Nuclear Polarisation (DNP) coupled SSNMR

DNP-SSNMR overcomes the intrinsic limitation of SSNMR, low sensitivity, and gives significant signal enhancement comparing to conventional SSNMR, which can dramatically

shorten the experiment time and allow signals to be observed for chemical species that would be too low in abundance to be observed in conventional SSNMR. In this work, it enables 2D heteronuclear correlation experiments between two dilute nuclei, e.g. ^{13}C-^{15}N, with significantly reduced experiment time and allows detection of minor ^{13}C species beyond the detection limit of conventional SSNMR.

DNP is achieved by transferring spin polarisation from unpaired electrons in so-called polarising agents to nearby nuclei in the sample of interest. The polarisation transfer is driven by microwave irradiation near the electron spin resonance frequency [129] [130]. The polarising agent used in this project is AMUPol [131], which has two unpaired electrons; its structure is shown in Figure 4-8. Typical DNP enhancement factors using AMUPol on in vitro ECM samples are between 20-60 on the 400 MHz SSNMR spectrometer used in this project. DNP experiments need to be carried out at very low temperature to enable the polarisation transfer and avoid unwanted side reactions involving the unpaired electrons. This low operating temperature, resulting in slow molecular motion, typically causes significant line broadening for biological samples.

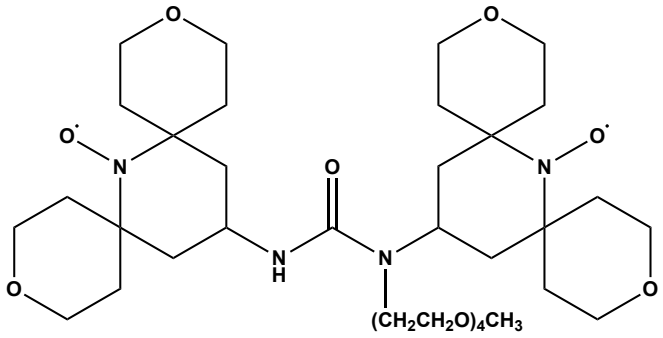


Figure 4-8 Structure of AMUPol, the polarising agent used in this project.

4.6 Two-Dimensional SSNMR Experiments

1D SSNMR spectra often lack the resolution needed for structure determination due to signal overlap, especially where signals are broad due to sample inhomogeneity as is the case for many samples in this work. 2D SSNMR experiments give more detailed structural information by resolving overlapping signals and correlating nuclei pairs according to their spatial proximity in the sample.

A typical 2D NMR experiment consists of four stages (Figure 4-9). It starts from a preparation period, usually a ^{1}H 90° pulse or ^{1}H-X CP, and then evolution. The evolution period, t_1, is not a fixed time period but incremented in a series of 1D experiments. Each 1D experiment, called a slice, is repeated with different increment delays, t, $t + \Delta t$, $t + 2\Delta t$, ⋯. This incrementation creates the second spectral dimension, called the indirect dimension or F1 dimension, in 2D NMR experiments. In practice, quadrature detection is achieved in the F1 dimension. The States-TPPI method is the most commonly used quadrature detection method in this project. Next, the mixing period is a period during which spins are allowed to communicate via different interactions. Interactions deployed in this period, mostly the dipolar interaction, determine what structural information is contained in 2D NMR spectra. More details about possible interactions and available structural information will be covered in later sections. Finally, signals are acquired during t_2.

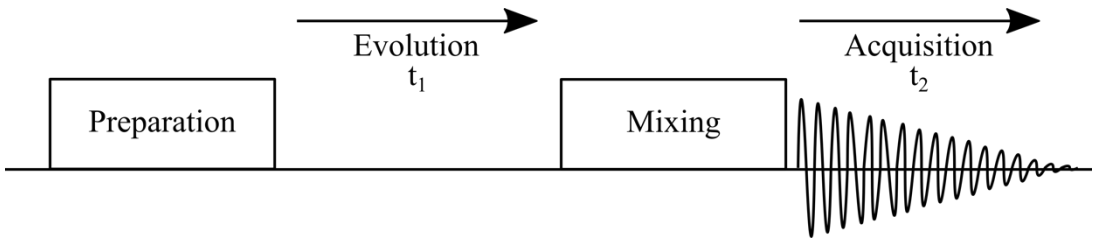


Figure 4-9 A general scheme for 2D SSNMR experiments, containing four stages, preparation, evolution, mixing and acquisition.

2D data acquired in 2D NMR experiments is Fourier transformed in both dimensions to generate 2D spectra as contour plots with chemical shifts on each dimension. In 2D NMR spectra, each signal has one chemical shift on each dimension, direct or F2 dimension and indirect or F1 dimension. Chemical shifts on F2 dimension are exact chemical shifts of observed nuclei, indicating their chemical environments; while chemical shifts on F1 dimension can mean different things depending on experiment types or interactions involved during the mixing period. Several frequently used 2D SSNMR MAS experiments will be introduced in more detail in the following sections, focusing mostly on 2D SSNMR spectra interpretation.

4.6.1 Homonuclear Correlation Experiments

^{13}C-^{13}C Proton-Driven Spin Diffusion (PDSD) and Dipolar-Assisted Rotational Resonance (DARR) Experiments

The dipolar interaction is distance-dependent and therefore can reveal structural information on internuclear distance. However, the dipolar interaction is averaged to zero by MAS, so this interaction needs to be reintroduced to study the ^{13}C-^{13}C dipolar interactions.

One of common experiments to reintroduce the dipolar coupling and correlate the spin pairs close in space is the 2D PDSD [132] and DARR [133] experiments. Their pulse sequences (Figure 4-10) begin with an initial CP step between ^{1}H and ^{13}C, and follow a decoupling pulse applied on ^{1}H channel while ^{13}C magnetisation evolves during t_1. Prior to the mixing period, a 90° pulse is applied on ^{13}C channel to prepare for spin diffusion. During the mixing time, the decoupling is switched off to allow spin diffusion involving ^{1}H-^{1}H and ^{1}H-^{13}C dipolar coupling. The mixing time determines the observable range of inter-^{13}C distances. At the end of the mixing period, a 90° pulse is applied on ^{13}C channel in order to acquire ^{13}C signals with ^{1}H decoupling.

The DARR experiment is a simple variant of the PDSD experiment. The only difference is that during the mixing time, a weak decoupling pulse is applied on ^{1}H channel. It has been shown that DARR experiments yield more intense cross peaks of long-distance spin pairs than PDSD experiments [134]; however, this effect has not been observed at our SSNMR spectrometer field strength and typical MAS rates used in this work.

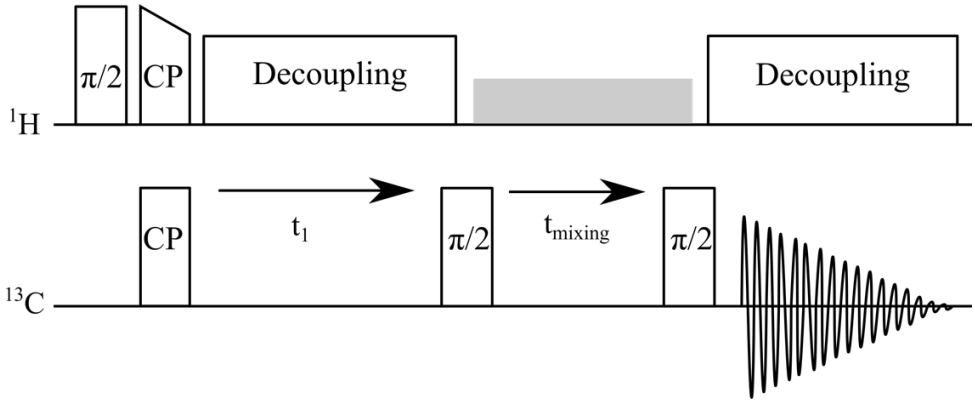


Figure 4-10 Pulse sequence for 2D ^{13}C-^{13}C PDSD and DARR experiments. During the mixing time, the ^{1}H decoupling (the grey rectangle) is switched off for PDSD experiments; for DARR experiments, the nutation frequency of this weak decoupling pulse is set to the MAS rate.

The 2D PDSD or DARR spectrum (Figure 4-11) is symmetrical about a diagonal (solid dark grey line in Figure 4-11) and has chemical shifts of the same nuclei, i.e. ^{13}C chemical shifts in this project, on both dimensions. Signals observed in 2D PDSD or DARR spectra are derived from dipolar-coupled nuclei pairs. Peaks on the diagonal are called diagonal peaks (bigger blue dots in Figure 4-11), indicating self-correlations between the same nuclei or environmentally similar nuclei, considering the broadness of peaks in SSNMR; peaks off the diagonal are called cross peaks (smaller blue dots in Figure 4-11), which originate from correlations between two environmentally different ^{13}C and thus give different chemical shifts on each dimension. A cross peak with chemical shift δ_1 on F1 dimension and δ_2 on F2 dimension means that a ^{13}C spin, $^{13}C_1$, giving chemical shift at δ_1 is dipolar-coupled to a ^{13}C spin, $^{13}C_2$, giving chemical shift at δ_2. In other words, $^{13}C_1$ and $^{13}C_2$ are dipolar-coupled ^{13}C pairs and close in space.

2D PDSD or DARR spectra allow identification of dipolar-coupled nuclei pairs, which can provide essential structural information, including chemical shifts and ^{13}C spatial proximity, to determine molecular structures. Previously, experiments have been performed by former group members, Dr. Wing Ying Chow and Dr. Veronica WC Wong, showing a mixing time of 20 ms at a MAS rate of 10 kHz, the commonly used experimental condition in this project, can detect ^{13}C correlating pairs up to 7 Å apart [135][136].

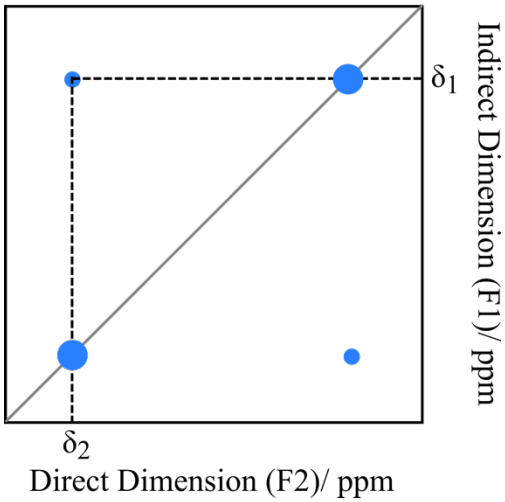


Figure 4-11 A schematic 2D PDSD or DARR spectrum showing two peaks on the diagonal and two cross peaks symmetrical to the diagonal. This spectrum shows that the nuclei giving rise to the chemical shift at δ_1 on F1 dimension are close in space to those giving the chemical shift at δ_2 on F2 dimension.

^{13}C-^{13}C Single Quantum-Double Quantum (SQ-DQ) Experiments

Another experiment to reintroduce the dipolar coupling under MAS is to generate double quantum magnetisation between dipolar-coupled spin pairs by symmetric homonuclear dipolar recoupling sequences (SHDRSs). SHDRSs are used to selectively recouple homonuclear dipolar-coupled spin pairs of interest, e.g. ^{13}C-^{13}C, so, unlike PDSD/DARR experiments, it is vital to avoid any magnetisation transfer from or to ^{1}H and requires a high level of ^{13}C isotope enrichment.

Several SHDRSs has been developed to date, such as C7 [137], POST-C7 [138] and SPC-5 [139]. They can be applied in 1D experiments, after CP and before acquisition, as a DQ filter to filter out natural abundant components of samples. This experiment will be referred to as 1D PC7 in the following texts. DQ filtration causes an intensity loss between normal CP and DQ-filtered CP. The extent of retaining signal intensity, or DQ efficiency, is used as a measurement of isotope enrichment level. For ^{13}C labelled ECM samples used in this project, DQ efficiency is around 20-30%.

Pulse sequence used in this project is a modified version of POST-C7 with large sweep width (LSW) on the DQ or indirect dimension [140] (Figure 4-12) to increase the spectral width on the DQ dimension to include the full chemical range. This 2D SQ-DQ LSW-POST-C7

experiment will be referred as 2D PC7 in the following text for simplification. After the initial ^{1}H-^{13}C CP step, a ^{13}C 90° pulse is applied to put the magnetisation back to the direction of B_0. Then, a rotor synchronised LSW-POST-C7 pulse train excites the magnetisation to double quantum and converts back to zero quantum after evolution period, t_1, for acquisition. Each LSW-POST-C7 pulse train contains seven elements, and they finish in 2 rotor periods. Then a final 90° pulse generates transverse magnetisation from the zero quantum magnetisation for observation.

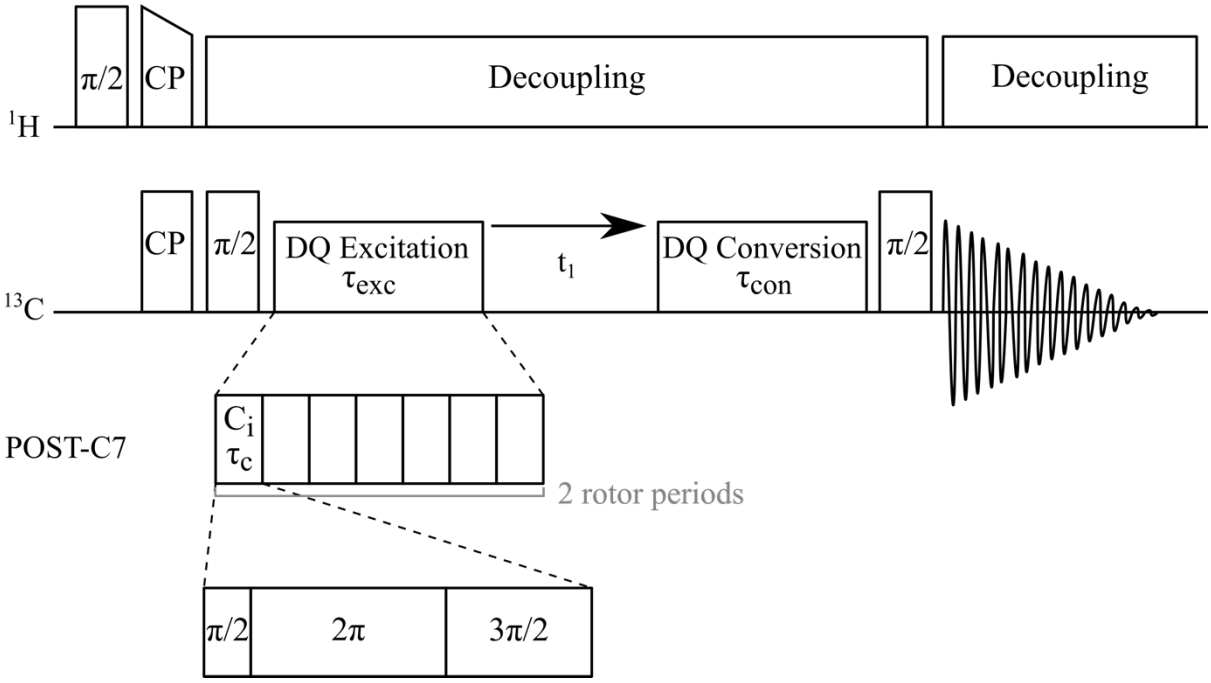


Figure 4-12 Pulse sequence for the 2D ^{13}C-^{13}C PC7 experiment. Above is a generic illustration of 2D SQ-DQ experiments. Specially designed SHDRSs are used to excite and convert magnetisation between SQ and DQ. POST-C7 sequence shown below is used in this project for SQ-DQ transitions. Each sequence contains 7 elements, C_i, and length of each element is τ_c. Each element contains 3 pulses in an order of 90°, 360° and 270°. The excitation period, τ_{exc} and the conversion period, τ_{rec} are equal and include a multiple of pulse trains, $\tau_{exc} = \tau_{rec} = n\tau_c$ $(n = 7, 14, \cdots)$. t_1 needs to be a multiple of τ_c, $t_1 = m\tau_c$ $(m = 0, 1, \cdots)$.

2D SQ-DQ experiments using SHDRSs, after Fourier transformation on both dimensions, give rise to 2D spectra with SQ chemical shifts on the F2 dimension and DQ chemical shifts on the F1 dimension, meaning signals are no longer symmetrical about a diagonal. A schematic 2D SQ-DQ spectrum is shown in Figure 4-13. Chemical shifts on those two axes give very

different information. If two signals share the same DQ chemical shift on F1 dimension, as (δ_1+δ_2) in Figure 4-13, which is also the sum of their SQ chemical shifts on F2 dimension, δ_1 and δ_2 separately, means the nuclei giving rise to these two signals are correlated by the pulse sequence and therefore are close in space.

Inter-^{13}C distances detected by 2D SQ-DQ experiments are relatively short. Experiments have also been performed by previous group members, Dr. Wing Ying Chow and Dr. Veronica WC Wong, showing 2D PC7 with n = 14 at a MAS rate of 10 kHz can detect ^{13}C correlating pairs 1-3 Å apart, roughly one bond distance away [135][136]. This is because the strength of dipolar coupling is highly dependent on inter-^{13}C distances, proportional to $\frac{1}{r^3}$. In other words, dipolar coupling between directly bonded ^{13}C pairs can be more than five times stronger than ^{13}C pairs with two bonds apart. This strong distance dependence makes this type experiment resemble scalar coupling mediated experiments, and correlating spin pairs observed in its spectra can be interpreted as directly bonded ^{13}C pairs [140]. This feature allows the establishment of ^{13}C bond connectivity within a molecule. For example in Figure 4-13, assuming it is a ^{13}C-^{13}C SQ-DQ spectrum, $^{13}C_1$ (δ_1) is close to $^{13}C_2$ (δ_2) because their DQ chemical shift equals the sum of their SQ chemical shifts, and $^{13}C_2$ (δ_2) is close to $^{13}C_3$ (δ_3). Therefore, it reveals ^{13}C connectivity information about this three-carbon molecule, $^{13}C_1$-$^{13}C_2$-$^{13}C_3$.

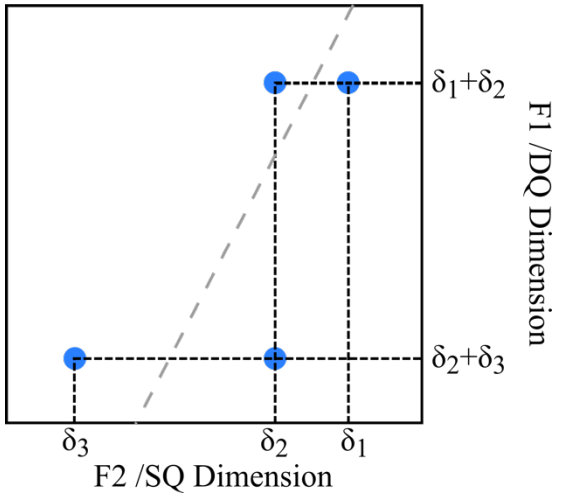


Figure 4-13 A typical 2D SQ-DQ spectrum, showing four signals. Signals on the same vertically dotted line share the same F2 chemical shift, meaning they are the same ^{13}C; while signals are horizontally connected and have a DQ chemical shift equal to the sum of their SQ chemical shifts meaning they are correlating with each other. Grey dash line is the DQ diagonal.

2D SQ-DQ experiments also separate correlation signals from self-correlated spin pairs or dipolar-coupled spins with similar chemical shifts, which would have been overlapped by the strong diagonal in 2D PDSD/DARR spectra.

SPC-5 sequences [139] used in DNP-SSNMR are variants of 2D PC7. They share the same principles but contain different number of elements in DQ magnetisation excitation and conversion sequence blocks. Interpretation of both 2D SQ-DQ SPC-5 and PC7 spectra is the same, and thus, the SPC-5 sequence will not be explained in detail.

4.6.2 Heteronuclear Correlation Experiments

^{1}H-^{13}C Heteronuclear Correlation (HETCOR) Experiments

The 2D ^{1}H-^{13}C HETCOR pulse sequence [141] used in this project can be seen as a 2D variant of CP sequence, shown in Figure 4-14. ^{1}H magnetisation is prepared by a 90° pulse. A t_1 evolution period is inserted prior to a CP contact pulse, during which decoupling pulses are applied on the ^{1}H channel to remove ^{1}H-^{1}H homonuclear interactions. ^{1}H-^{1}H homonuclear decoupling ensures the resolution on ^{1}H or indirect dimension. Also, shorter contact time, than normal CP, is used to limit ^{1}H-^{1}H interactions within that period. Varying contact time changes the observable range of ^{1}H-^{13}C distances.

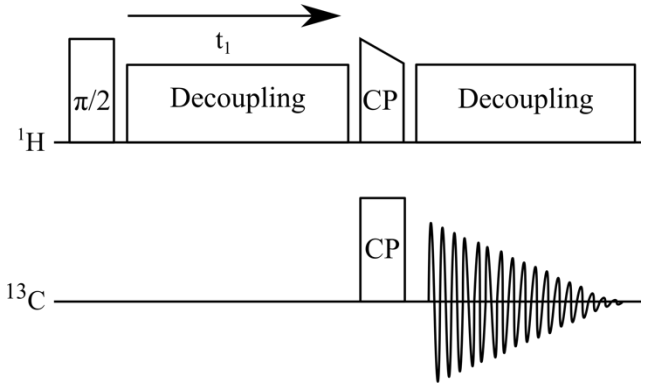


Figure 4-14 2D ^{1}H-^{13}C HETCOR pulse sequence. It begins from turning ^{1}H magnetisation 90° to prepare it for evolution in t_1. Afterwards, short CP pulses are applied on both channels for the final acquisition on ^{13}C with ^{1}H decoupling.

High resolution in the ^{1}H dimension is required in this experiment to extract ^{1}H chemical shifts, which is achieved by employing ^{1}H-^{1}H Frequency-Switched Lee-Goldburg (FSLG) decoupling sequence [142] on ^{1}H channel during the evolution period. FSLG decoupling sequence contains

a series of 360° pulses with alternating frequency to limit ^{1}H-^{1}H homonuclear dipolar interactions but allow ^{1}H-^{13}C heteronuclear dipolar interactions.

2D ^{1}H-^{13}C HETCOR experiments are achieved by CP, making it difficult to detect ^{1}H-^{13}C correlations in mobile environments. Free ^{1}H without any stabilisation factor, e.g., hydrogen bonds, is not likely to be detected by ^{1}H-^{13}C HETCOR parameters used in this project.

^{13}C-^{15}N Double Cross Polarisation (DCP) and DARR assisted DCP Experiments

Detecting correlations between ^{13}C-^{15}N is more challenging and time consuming than ^{1}H-^{13}C and ^{13}C-^{13}C correlations because of the small gyromagnetic ratio of ^{15}N. It also requires a high level of both ^{13}C and ^{15}N enrichment in samples. So 2D ^{13}C-^{15}N correlation experiments are only performed using DNP-SSNMR spectroscopy. Usually, 2D ^{13}C-^{15}N correlation experiments begin with ^{1}H-^{15}N/^{13}C CP, and then allow ^{15}N spins to interact with ^{13}C spins via dipolar interactions. Different pulse sequences have been designed to measure selective or non-selective ^{13}C-^{15}N interactions in biological samples.

A simple form of 2D ^{13}C-^{15}N correlation experiments is double cross polarisation (DCP), whose sequences (Figure 4-15) begin from magnetisation transfer from ^{1}H to ^{15}N and then from ^{15}N to ^{13}C and eventually acquire on ^{13}C channel [143][144]. However, the ^{13}C-^{15}N internuclear distance detected in DCP experiments is limited because the dipolar interaction is distance-dependent (explained in section 4.2.3). So, sometimes ^{13}C-^{15}N CP is followed by ^{13}C-^{13}C DARR mixing to obtain ^{13}C-^{15}N correlations with multiple bonds apart, as in DARR assisted ^{13}C-^{15}N DCP experiments. Its pulse sequence is shown in Figure 4-15. The inter-^{13}C-^{15}N distance detected in ^{13}C-^{15}N DCP DARR experiments is primarily determined by the DARR mixing time.

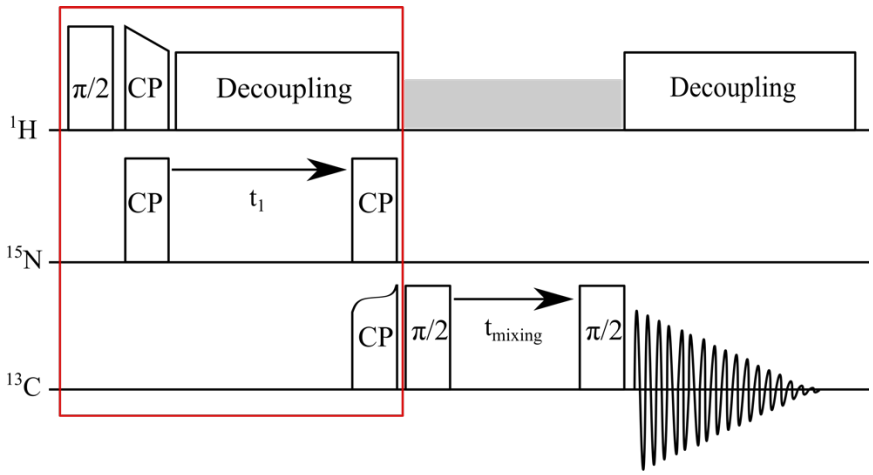


Figure 4-15 The pulse sequence of the 2D ^{13}C-^{15}N DCP followed by DARR mixing. The part included in the red rectangle is a simple DCP sequence, which starts from a CP between ^{1}H-^{15}N and is followed by a second CP between ^{15}N and ^{13}C. The ^{13}C-^{13}C DARR sequence has been explained in section 4.6.1.

Heteronuclear Correlation Spectra Interpretation

All the above-mentioned 2D heteronuclear correlation experiments generate 2D spectra with two dimensions showing chemical shifts of different nuclei separately, and therefore 2D heteronuclear spectra are not symmetric about a diagonal as some homonuclear correlation spectra. A schematic 2D heteronuclear spectrum is shown in Figure 4-16 with ^{13}C chemical shifts on the direct dimension (the horizontal axis in the figure) and X, ^{1}H or ^{15}N, on the indirect dimension (the vertical axis in the figure).

In addition, all heteronuclear correlation experiments described here are mediated via dipolar coupling, meaning that they can be interpreted in the same way, to some extent. For example, if the 2D spectrum in Figure 4-16 is acquired from a 2D ^{1}H-^{13}C HETCOR experiment, the vertical axis will show ^{1}H chemical shifts; and the correlation signal between $^{13}C_1$ and δ_2 means that ^{13}C spins giving a chemical shift at $^{13}C_1$ is dipolar coupled to ^{1}H spins giving a chemical shift at δ_2, and they are close in space. If the 2D spectrum in Figure 4-16 is acquired from a 2D ^{13}C-^{15}N correlation experiment, the vertical axis will show ^{15}N chemical shifts; and the correlation signal between $^{13}C_1$ and δ_2 means that ^{13}C spins giving a chemical shift at $^{13}C_1$ is dipolar coupled to ^{15}N spins giving a chemical shift at δ_2, and they are close in space.

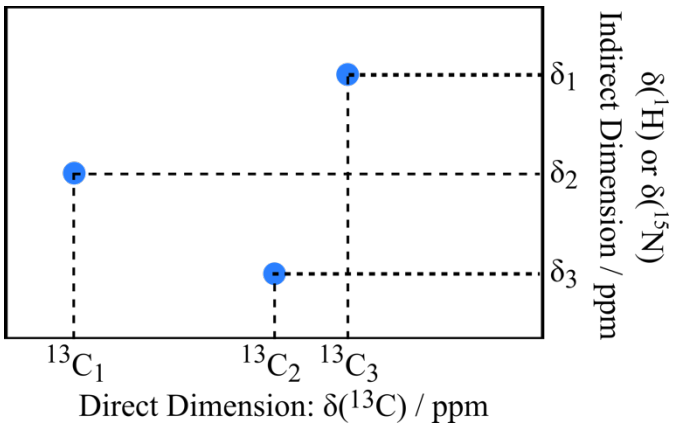


Figure 4-16 A schematic 2D ^{13}C-X heteronuclear correlation spectrum, showing three correlating ^{13}C-X pairs (indicated by blue dots). All above-mentioned heteronuclear experiments acquire on ^{13}C channel, so the direct dimension (horizontal axis) shows ^{13}C chemical shifts and the indirect dimension (vertical axis) shows X chemical shifts. In 2D ^{1}H-^{13}C HETCOR experiments, the indirect dimension will be ^{1}H chemical shifts and in 2D ^{13}C-^{15}N correlation experiments, the indirect dimension will show ^{15}N chemical shifts.

5 Experimental Details

5.1 In vitro Extracellular Matrix Preparation

Cells were cultured in a T-175 flask containing 25 mL complete Dulbecco's Modified Eagle Medium (DMEM with 1 g/L glucose; Invitrogen). Complete DMEM was composed of 10% foetal calf serum (First Link), 30 μg/mL L-ascorbic acid 2-phosphate (Sigma) and 10 mL/L L-glutamine-penicillinstreptomycin (200 mM L-glutamine, 10,000 units/ml penicillin, and 10 mg/ml streptomycin in 0.9% sodium chloride; Sigma). Cell culture flasks were incubated at 37°C in a humidified atmosphere of 95% air and 5% CO_2, and the culture medium was renewed every 2 days.

L-ascorbic acid 2-phosphate solution was prepared in phosphate buffered saline (1X PBS, Invitrogen) from powder and added in cell culture medium after filter sterilisation (0.22 μm syringe filter, Appleton Woods).

ECM was harvested when cells produced a dense matrix which started to peel off the surface of the tissue culture flask. This usually takes about 10-15 days from seeding cells. Medium was removed, and cells were washed twice with 10 mL PBS to remove the cell culture media residual before any decellularization procedure.

Freeze-thaw lysis

The flask was placed in a freezer at -80°C for overnight and was allowed to thaw at room temperature for 30 minutes. The ECM was dislodged by gently swirling the flask in the presence of PBS. The ECM collected in PBS was transferred to a fresh 50 mL centrifuge tube. The ECM was further washed repeatedly by PBS to remove cell debris. Samples were stored at -20°C prior to any further procedure.

Detergent lysis

10 mL of working detergent solution (0.5% Triton X-100, 20 mM NH_4OH in 1x PBS) was added into each T-175 cell culture flask. The flask was incubated at 37°C for 10 mins. The decellularised ECM was dislodged by gently swirling the flask in the presence of detergent working solution. Then it was collected and transferred to a 50 mL centrifuge tube containing 10 mL PBS. This tube was left in a 4°C fridge for overnight.

The ECM was washed by PBS three times to remove remaining detergent, and was incubated with DNAse (Thermo Fisher Scientific, 10 ug/mL, 1:1000 dilution in its reaction buffer) at 37°C for 30mins to remove residual DNA left by cells. DNAse solution was aspirated and matrix was thoroughly washed before transferring to a fresh 50 mL centrifuge tube. Samples were stored at -20°C prior to any further procedure.

Introducing Isotope Labelling

When isotope labelled matrix was needed, stock isotope labelled amino acids solution was prepared in PBS from powder and filter sterilised (0.22 μm syringe filter, Appleton Woods). U-^{13}C, ^{15}N-amino acids (Cambridge Isotope Laboratories) solution was added in cell culture medium to a final concentration of double the initial concentration in the DMEM, after cells were confluent. They were refreshed every 2 days together with cell culture medium until ECM is ready to harvest.

Table 5-1 Concentrations of U-^{13}C, ^{15}N-amino acids in cell culture medium.

Amino Acid	Concentration /mM
L-Arg	0.8
L-Lys	1.6
L-Pro	0.8
Gly	0.8

5.2 Collagen Extraction from In vitro Extracellular Matrix

To remove minor matrix proteins and extract collagen from in vitro extracellular matrix, one flask of freshly decellularised ECM was incubated with chymotrypsin (0.1 mg, Sigma, in 2 mL buffer, 100 mM Tris-HCl, 10 mM $CaCl_2$, pH=7.8 in water) at 37°C for overnight and washed. Then it was incubated with Tween-20 solution (0.25% in 2 mL PBS) at 37°C for another overnight and washed. The sample was stored at -20°C in a centrifuge tube until further procedure.

5.2.1 Transmission Electron Microscopy (TEM) Imaging

ECM samples are gentled sonicated in water using a probe sonicator at a power about 10 W for 5-10 seconds, depending on the amount of the ECM sample, to better disperse sample in solution. Sonication can be repeated if the ECM sample is not well-dispersed. 5 μl of solution was adsorbed onto glow-discharged 400 mesh copper/carbon-film grids (EM Resolutions) for about 2 min. Grids were rinsed on two drops of DIW and negative staining was performed using a 2 % aqueous uranyl acetate solution. Grids were viewed in a FEI Tecnai G^2 electron microscope run at 200 keV using a 10 μm objective aperture. Images were acquired using AMT Camera software.

5.2.2 Fluorescence Spectroscopy

All the fluorescence spectra were recorded on a Cary Eclipse fluorescence spectrophotometer (Agilent, former Varian). ECM was gently slided in a micro quartz cuvette with a pathlength of 10x2 mm and a chamber volume of 700 μL (Hellma Analytics). Voltage typically used were between 600 W – 750 W depending on sample volume. To acquire a spectrum with decent signal to noise, higher voltage was usually used when sample volume was small.

Table 5-2 Excitation wavelength used, and emission spectra width acquired in this project.

Excitation wavelength / nm	Emission wavelength range / nm
275	285-450
285	295-450

5.2.3 Amino acid analysis

Amino acid analysis service was provided by the Department of Biochemistry, University of Cambridge. Freeze-dried ECM samples were provided to the service for analysis.

The procedure is modified from reference [145] for the ECM samples. Briefly, freeze-dried ECM samples were hydrolysed by a mixture of concentrated HCl, phenol (0.5 mL) and dodecanthiol (0.86 μL), with the presence of known amount of L-norleucine as an internal standard, under vacuum at 115 ℃ for 22 hours. Then the hydrolysed sample was dried and neutralised in a desiccator with solid NaOH and re-dissolved in buffer (sodium citrate, pH = 2.2, Sigma) before injecting to a Pharmacia Alpha Plus series amino acid analyser (Biochrom Ltd, Cambridge, UK). Chromatography was performed and peak detection was achieved by mixing the eluate with ninhydrin at 135 ℃ and measuring the absorbance at 570 and 440 nm. The peak area from each amino acid was recorded and compared with the internal standard to give the ultimate mol% for each amino acid.

5.3 Glycation

Collagenous samples, either intact ECMs or collagen extracted from ECMs were incubated in 50 mM U-$^{13}C_5$-Rib or -R5P or 100 mM U-$^{13}C_6$-Glc in 150 mM phosphate buffer at 37 ℃ for 8 weeks. Often, ECM harvested from 2-3 flasks of cells were incubated in 6 mL solution. At the end of the glycation incubation, the collagenous sample was washed with deionised water to remove any excess glycating agents and phosphate buffer. The sample was freeze-dried for overnight and stored in -20 ℃ freezer before further analysis.

5.4 Sample Preparation for SSNMR Experiments

All SSNMR samples were thoroughly washed by water to remove possible salt content from PBS. Then they were freeze-dried overnight and stored at -20°C before SSNMR experiments.

SSNMR

Samples were packed in Kel-F inserts that fit in 4 mm zirconia rotors, unless they have sufficient amount to be packed directly in a rotor.

Generally speaking, sample mass varies between different sample types. Most samples included in this thesis weigh around 10-20 mg. Low sample mass can lead to low S/N ratio and thus less informative/desirable spectra.

DNP-SSNMR

For the DNP NMR experiments, 3.2 mm outer diameter zirconia rotors with Vespel caps were used. AMUPol was prepared as 20 mM solutions in D_2O/H_2O (75% D_2O, 25% H_2O), then 12.5 µL of the radical solution was added to the sample-filled rotor. The rotors were closed with a Vespel cap, then spun in both vertical orientations (with the cap facing upwards and downwards) with a desktop mini centrifuge to distribute the radical solution within the rotor. The packed rotors were allowed to equilibrate with the radical solution overnight in the fridge before the start of DNP NMR experiments.

5.5 SSNMR Experimental Parameters

5.5.1 Spectrometer Used, MAS Rates and Chemical Shift Referencing

SSNMR

All experiments were performed on a Bruker AVANCE I 400 MHz wide bore spectrometer operating at a static magnetic field of 9.4 T, with ^{31}P Larmor frequencies at 162 MHz, ^{13}C at 100 MHz and ^{15}N at 40 MHz. A standard 4mm double resonance probe with a MAS rate limit of 15 kHz was used for all experiments.

Chemical shifts reported in this thesis were referenced on different chemicals, as shown in the table below (Table 5-3), as well as MAS rates used to acquire different nuclei. All 2D spectra were acquired at a MAS rate of 10 kHz. Most experiments were performed under room temperature without temperature control/chilling; some were at 290 K.

Table 5-3 Chemical shift references and MAS rates for acquiring different nuclei.

Nuclei	MAS speeds	Chemical Shift References
^{31}P	10 kHz	Hydroxyapatite at 2.6 ppm with respect to 85 wt% H_3PO_4 at 0 ppm
^{13}C	10 kHz	Glycine C_α at 43.1 ppm with respect to TMS at 0 ppm
^{15}N	6 kHz	Glycine ^{15}N at 32.3 ppm with respect to ammonia at 0 ppm
^{1}H	13 kHz	Ethanol methyl carbon at 1.11 ppm with respect to TMS at 0 ppm

DNP-SSNMR

DNP-SSNMR experiments were performed on a Bruker Avance III wide-bore spectrometer operating at a static magnetic field of 9.4 T, with ^{1}H Larmor frequencies of 400 MHz. For DNP-SSNMR, the 400 MHz spectrometer is equipped with a gyrotron operating at 9.7 T and a 3.2 mm (^{1}H- ^{13}C-^{15}N) triple resonance low-temperature magic angle spinning (LTMAS) probe. DNP-enhanced experiments were carried out under 30 mA microwave irradiation.

All spectra were acquired at a MAS rate of 8889 Hz. The sample temperature was set to 100 K using the LTMAS cooling cabinet.

5.5.2 Typical Parameters for SSNMR Experiments

Typical CPMAS SSNMR Experiments Parameters

In this project, CPMAS SSNMR experimental parameters are usually optimised on solid-state collagen-like peptides, such as $(GPO)_{10}$. Only if necessary, parameters are further optimised on individual samples. Typical ^{1}H-X CPMAS parameters for this work are included in Table 5-4. ^{1}H's natural abundance is 99.99%.

Table 5-4 Typical ^{1}H-X CPMAS SSNMR experiments parameters used in this project to observe different nuclei. They are all obtained collagen-like peptides with a MAS rate at 10 kHz.

Nuclei observed	**Natural abundance**	**Recycle delay**	**^{1}H 90^{0} pulse length**	**Contact pulse length**	**Decoupling**
^{13}C	1.11%	2 s	2.5 μs	2.5 ms	104 kHz SPINAL64
^{15}N	0.36%			2.5 ms	
^{31}P	100%			10 ms	

Typical CPMAS DNP-SSNMR Experiments Parameters

DNP-SSNMR parameters on the 400 MHz spectrometer were optimised on the specific sample but they usually are in a similar range: ^{1}H 90° pulse length 2.2-2.5 μs, 1500 μs ^{1}H- ^{13}C ramped CP contact time, 80+ kHz ^{1}H decoupling, with a recycle delay 1.5 s.

Typical 2D ^{13}C-^{13}C PDSD MAS SSNMR Experiments Parameters

^{1}H-^{13}C CP parameters have been included in the previous sections. ^{13}C 90° pulse length is 3.57 μs. 104 kHz CW-13 is used during the incremental delay and 104 kHz SPINAL64 decoupling scheme is used during acquisition periods.

Typical 2D ^{13}C-^{13}C DARR MAS DNP-SSNMR Experiments Parameters

^{1}H-^{13}C CP parameters under DNP conditions have been included in the previous sections. ^{13}C 90° pulse length is usually between 4.1-4.75 μs. ^{1}H decoupling pluses are typically between 80-100 kHz.

Typical 2D ^{13}C-^{13}C PC7 MAS SSNMR Experiments Parameters

^{1}H-^{13}C CP parameters have been included in the previous sections. At our experimental condition, MAS rate of 10 kHz, ^{13}C 90° pulse length is set to 3.57 μs to achieve the requirement on field strength, $\omega_{rf} = 7\omega_r = 70\,kHz$. t_1 is set to be $28.57\,\mu s = \tau_c$. With n=14, $\tau_{exc} = \tau_{rec} = 0.4\,ms$. 104 kHz LG decoupling scheme [146] is used during C7, and 104 kHz SPINAL64 decoupling scheme is used during acquisition periods. LG decoupling scheme

introduces a scaling factor on ^{1}H dimension, hampering ^{1}H chemical shifts. This is taken care of post-acquisition.

Typical 2D ^{13}C-^{13}C SPC-5 MAS DNP-SSNMR Experiments Parameters

^{13}C-^{13}C 2D single quantum-double quantum (SQ-DQ) correlation spectra were carried out by ^{1}H-^{13}C CP followed by a SPC-5 mixing sequence. The MAS rate was set to 8889 Hz. DQ coherence was excited by applying the SPC5 sequence for 1.2 ms (10 times).

Typical 2D ^{1}H-^{13}C HETCOR MAS SSNMR Experiments Parameters

Correlations between ^{1}H and ^{13}C are investigated using HETCOR experiments. Typical parameters are listed in the table below.

Table 5-5 Typical parameters used in 2D ^{1}H-^{13}C HETCOR MAS SSNMR experiments for observing different nuclei. They are all obtained with a MAS rate of 10 kHz.

Nuclei observed	**Recycle delay**	**^{1}H 90° pulse length**	**Contact pulse length**	**^{1}H-^{1}H Decoupling (evolution)**	**^{1}H-^{13}C Decoupling (acquisition)**
^{13}C	2 s	2.5 μs	0.5 ms	104 kHz FSLG	104 kHz SPINAL-64

Typical 2D ^{13}C-^{15}N Correlation DNP-SSNMR Experiments Parameters

2D ^{13}C-^{15}N heteronuclear correlation experiments were carried out with 90° pulse lengths of: ^{1}H 2.4 μs, ^{13}C 4.33 μs, ^{15}N 4.27 μs. In ^{13}C-^{15}N DCP-DARR experiments, magnetization was first excited by ^{1}H-^{15}N CP with a 450 μs contact pulse length, with a slight ramp shape on the ^{15}N channel, followed by indirect ^{15}N evolution, during which ^{1}H was decoupled at 80 kHz. Magnetization was then transferred by ^{15}N-^{13}C SPECIFIC-CP with a 4.5 ms contact pulse length, with a square shape on ^{15}N and a tangential shape on ^{13}C. At this point there was no indirect evolution on ^{13}C, instead, a DARR sequence was immediately applied on the ^{13}C channel, followed by ^{13}C acquisition with 80 kHz ^{1}H decoupling. A 219 ppm ^{15}N sweep width was used here to ensure the full ^{15}N range was recorded without aliasing. 60 indirect evolution increments were collected, with 1024 transients per increment. Thus, the signal averaging time was ~51 hours.

5.6 Cellular Metabolites Extraction

Cells were cultured in a T-175 flask containing 25 mL Dulbecco's Modified Eagle Medium (DMEM; Invitrogen) with 10% foetal calf serum (First Link) and 10 mL/L L-glutamine-penicillinstreptomycin (200 mM L-glutamine, 10,000 units/ml penicillin, and 10 mg/ml streptomycin in 0.9% sodium chloride; Sigma). For K7M2, DMEM with 4.5 g/L glucose was used; while for VSMC, DMEM with 1 g/L glucose was used. The culture was incubated at 37 °C in a humidified atmosphere of 95% air and 5% CO_2, and the culture medium was renewed every 2 days until cells reached 90-95% confluency. Then medium was removed, and the cells were washed once with 10 mL saline (Invitrogen).

Afterwards, cells were lysed by either freeze-thaw lysis or liquid nitrogen lysis:

Freeze thaw lysis (FT)

Cell culture flasks were placed in -80°C freezer for overnight and thawed the next day under room temperature for 30 minutes. Initially, cells were scraped off with the presence of 5-8 mL water; later, cells were scraped off with the presence of cold 80% MeOH (8 mL, pre-cooled in -80 °C freezer) to extract metabolites. Then, the mixture was transferred into a fresh centrifuge tube and it was centrifuged at 4 °C, 4500 RPM for 15 mins. Cell debris was washed again with cold 80% MeOH (8 mL, pre-cooled in -80 °C freezer) to achieve better extraction.

Liquid nitrogen lysis (LN)

Liquid nitrogen was poured into cell culture flasks to freeze cells. Pre-cooled 80% MeOH (8 mL, -80 °C) was added to a flask to extract cell metabolites. Cells were scraped off with the presence of cold 80% MeOH, and transferred to a fresh centrifuge tube. Later, cell mixture was gently sonicated in ice to break more membranes. The tube was vortexed for better extraction, and centrifuged down at 4500 RPM for 15 minutes. Cell debris was washed again with cold 80% MeOH (8 mL, pre-cooled in -80 °C freezer) to achieve better extraction.

Finally, clear supernatant was transferred into 50 mL centrifuge tube and freeze-dried overnight. Freeze-dried cell lysate was stored at -20 °C freezer before further experiments.

5.6.1 Poly-L-Lysine Incubation Experiments

Freeze-dried cell lysate samples were reconstituted in 1.4 mL 150 mM phosphate buffer (PB). The mixture was then centrifuged at 4500 RPM at 2 °C for 5 minutes to remove undissolved precipitates. Poly-L-Lysine (PLL, Sigma) was dissolve in PB to make 1 mL of 100 mM PLL solution. Same with glucose and other sugar, 1 mL 100 mM solution was made in PB. 100 μL PLL and 100 μL reconstituted cell lysate were added into a well in a black 96-well-plate for fluorescence measurement. 5 replicates were made for each experimental group. Control groups were made by adding 100 μL PB together with 100 μL control samples to balance the concentration. Plates were incubated in a 37 °C incubator, and AGE fluoresecence was measured at different time points over incubation.

AGE Fluorescence Measurement

AGE fluorescence was measured using a FLUOstar Omega plate reader (BMG LABTECH). Fluorescence intensities were recorded at excitation 355 nm and emission 460 nm.

Data processing: Within the same plate, data acquired at different time points are normalised on PLL, which is not supposed to be fluorescent; however always give a fluorescent reading around 10 times of PBS-noise level intensity.

5.6.2 Solution-state NMR experiments

700 μL D_2O with TSP as internal NMR reference, was added into lyophilized cell lysate (from 1 flask of cells). It was then centrifuged at 4500 RPM at 2 °C for 3 minutes to remove bubbles and precipitates. Clear supernatant was transferred into a 5mm solution state NMR tube.

All the solution-state NMR spectra were acquired on a 500 MHz Bruker spectrometer equipped with a cryoprobe that can operate at 5 ℃.

5.6.3 HPLC-UV-MS Experiments

Pre-column Derivatisation

Derivatisation method was adapted from reference [147].

Different reaction times, 1, 2, 3, 4 and 9 hours, were tested on 0.5 mg/mL R5P in 75% methanol in water with an internal standard solution of 0.2 mg/mL fucose in in 75% methanol in water at 70 °C, static. Optimal reaction time was found to be 1 hour.

A calibration curve was established with R5P in 75% methanol in water of different concentrations: 1 mg/mL, 0.5 mg/mL, 0.25 mg/mL, 0.1 mg/mL and 0.05 mg/mL, with the same concentration of fucose as internal standard.

100 uL 75% methanol in water was added in lyophilized cell lysate to dissolve as much as possible, and was transferred into a 1.5 mL tube. The tube was centrifuged at 10,000 rpm for 5 mins, and the supernatant was taken out as sample A. Then, a further 100 uL 75% methanol in water was added in the remaining cell lysate to repeat the process. The second supernatant was regarded as sample B.

50 uL reconstituted cell lysate was mixed with 50 uL internal standard solution (0.2 mg/mL fucose in 75% methanol in water) in a 2 mL Eppendorf tube, followed by the sequential addition of 100 uL of 25 mM AEC in methanol, 50 uL of 50 mM $NaCNBH_3$ solution, and 20 uL acetic acid. The mixture was heated at 70 °C for 1 hour in a heating block. After reaction, the tube was cooled on ice for 1 min, and then, 100 uL of water and 1 mL of a dichloromethane-hexane (2:1, v/v) solvent were added. Each tube was vortexed and was then centrifuged at 10,000 rpm for 5 mins in a microcentrifuge. A 100 uL aliquot of the upper aqueous phase was transferred to an LC autosampler vial for HPLC-UV-MS analysis.

HPLC-UV-MS Detection and Quantification

All LC separations were carried out using Thermofisher HPLC system with a Phenomenex Synergi C18 (250×4.6mm, 4μm, 80Å) column. The mobile phase was composed of water/0.1% formic acid (A) and acetonitrile/0.1% formic acid (B). The column flow rate was set at 1

mL/min, and the column temperature was maintained at 30 ℃. The elution gradient was 0% to 35% B in 45 mins, to 100% B in another 5 mins, and then held at 100% for 5 mins before equilibrating at 100% A for a further 5 mins. UV absorption spectra were collected at 254 nm.

6 Collagen Extraction from In vitro ECM

In vitro ECM is known to contain hundreds of proteins other than collagen (described in section 2.3). Extracting collagen from uniformly isotope enriched in vitro ECM allows the production of uniformly isotope labelled collagen, which enables further studies of collagen and collagen glycation by SSNMR.

However, removing non-collagenous (NC) ECM proteins and extracting intact collagen from in vitro ECM can be challenging, because commonly used proteolytic enzymes, such as trypsin, papain and pepsin, would digest collagen at the same time as NC proteins [148][149]. An ideal enzyme for collagen extraction from in vitro ECM requires (1) working best under physiological conditions, pH and temperature, to avoid the effects of extreme working conditions on collagen and (2) working exclusively or at least significantly more effectively on NC ECM proteins than collagen, in other words, preferable enzymatic cleave sites involving amino acids more abundant in minor proteins, less abundant or even absent in collagen.

Chymotrypsin was eventually chosen, aiming to remove NC ECM proteins while keeping collagen intact. Its activity is maximal at pH=7.8-8, but it is effective enough to use at physiological pH=7.4 [150]. This gentle pH condition minimises the possibility of collagen hydrolysis under acidic conditions. Moreover, chymotrypsin cleaves preferentially at carbonyl side of amide bonds of large hydrophobic or aromatic amino acids, such as phenylalanine (F), tyrosine (Y), tryptophan (W) and leucine (L), unless the next amino acid is proline (P) [151][152]. Considering the amino acid composition of collagen versus other ECM proteins, there are far fewer aromatic amino acids in collagen, and thus providing far fewer cleavage sites. For example, in bovine collagen type I, about 2.3% of the amide bonds are theoretically cleavable, while in bovine fibronectin isoform 1 this number goes up to 12.6% (Table 6-1) [153]. Additionally, due to the complex organisation of ECM, collagen may not be directly exposed to surrounding enzymes, and therefore can be regarded as protected to some extent by collagen binding proteins. Thus, collagen may not be digested until these binding proteins are removed.

Thus, chymotrypsin appears an ideal enzyme to remove minor ECM proteins and extract intact collagen fibrils from in vitro ECM. Its candidacy was tested by different characterisation methods, designed to (1) investigate effective working conditions on two different types of ECMs, as composition of ECM from different cells can be different, to see if a generic and robust protocol can be established (2) evaluate the effectiveness of the protocol, if successfully established, and (3) examine the purity and intactness of the resultant extracted collagen fibrils. Results and experimental details will be presented in the following sections.

Table 6-1 Number of cleavable F, Y, W and L sites in Top: Bovine Collagen Type I. Middle: Bovine Fibronectin Isoform 1. Bottom: Bovine Elastin Isoform 1. Numbers of residues are calculated using sequences published on protein sequence database, UniProt [153]. Identifiers for Bovine Collagen Type I are P02453 and P02465, for Bovine Fibronectin Isoform 1 is P07589-1, for Bovine Elastin Isoform 1 is P04985-1. Numbers of residues can vary from one isoform to another resulting from slightly different sequences between isoforms. For ease of demonstration, the calculation below assumes that chymotrypsin only works on the 4 amino acids mentioned here.

Amino acid (percentage per molecule)	Number in **Collagen I** α1 chain			Number in **Collagen I** α2 chain		
	Total	Protected by P	Digestible	Total	Protected by P	Digestible
F (1.3%)	13	7	5	16	5	11
Y (0.4%)	5	0	5	3	1	2
W (0)	0	0	0	0	0	0
L (2.5%)	22	12	10	34	15	19
			20 x2 =40			32
Sum						72 (2.3%)

Amino acid (percentage per molecule)	Number in **Fibronectin**		
	Total	Protected by P	Digestible
F (2.1%)	51	7	44
Y (4.1%)	101	4	97
W (1.6%)	40	0	40
L (5.8%)	143	12	131
Sum			312 (12.6%)

Amino acid (percentage per molecule)	Number in **Elastin**		
	Total	Protected by P	Digestible
F (2.8%)	21	6	15
Y (0.94%)	7	3	4
W (0)	0	0	0
L (6.7%)	50	12	38
Sum			57 (7.6%)

6.1 Effects of Chymotrypsin Digestion on In vitro ECM

In vitro ECM was decellularised and incubated with chymotrypsin; four techniques were used to examine the efficacy of this extraction: TEM, fluorescence microscopy, amino acid analysis (AAA) and SSNMR microscopy.

The decellularisation method is an important factor in determining ECM composition, therefore it was optimised at the beginning to obtain relatively "clean" ECM with as few as possible NC components [154][155]. Initially, ECM is subjected to freeze-thaw lysis here to lyse the cells, which preserves essentially all proteins as well as a significant amount of lipid. Then, a mild detergent is used to decellularize ECM by gently breaking lipid-lipid interactions and protein-lipid interactions, obtaining cleaner ECM than that obtained from freeze-thaw lysis.

6.1.1 Transmission Electron Microscopy (TEM)

TEM can visually and qualitatively demonstrate collagen quality and residual ECM components in differently processed ECMs. Collagen fibrils, under TEM, are long (typically longer than the field of view), relatively straight, i.e. not kinked, and have distinctive banding patterns (illustration in Figure 2-4). These unique characteristics of collagen fibrils enable accurate recognition of collagen fibrils and distinguish collagen from other ECM proteins in TEM images. Specifically, in this work, TEM is used firstly, to examine if chymotrypsin is able to remove NC proteins, and secondly, to decide an optimal chymotrypsin treatment condition for ECM that removes the most NC proteins and preserves collagen fibrils. In other words, TEM is used to assess the feasibility and effectiveness of using chymotrypsin to clean in vitro ECM.

To start with, ECM without any post-cell lysis treatment was observed under TEM. A representative image of freshly decellularised ECM from in vitro cell culture of bovine vascular smooth muscle cells (VSMCs) is shown in Figure 6-1 A. Different decellularisation methods do not show significant differences; both methods, freeze-thaw lysis and detergent lysis, show a substantial amount of what appears to be small proteins. Some of them are adhered to collagen fibrils while some are spread in the background. Collagen fibrils can be recognised but are mostly buried underneath small ECM proteins. Fibril banding patterns cannot be observed.

Then, different chymotrypsin incubation times were tested on VSMC ECM, 30 mins, 2 hrs and overnight (between 18-24 hrs) at 37℃ (detailed procedure is included in section 0), and eventually, overnight incubation was decided to be optimal for producing clean, well-structured collagen fibrils after comparing a considerable number of images obtained from different chymotrypsin incubation times (images are not shown here due to space limitation).

After overnight incubation (Figure 6-1 B), significant protein removal is achieved compared to untreated ECMs (Figure 6-1 A), and collagen banding patterns begin to be observed. This indicates that NC proteins are removed from collagen surface, revealing better details of the collagen fibril itself, though there is still some extraneous material left either associated with or dissociated from collagen fibrils. Overall, collagen fibrils are observed with a negligible amount of proteins entangled after overnight chymotrypsin incubation. Thorough washings should be done afterwards to remove unbound proteins as much as possible.

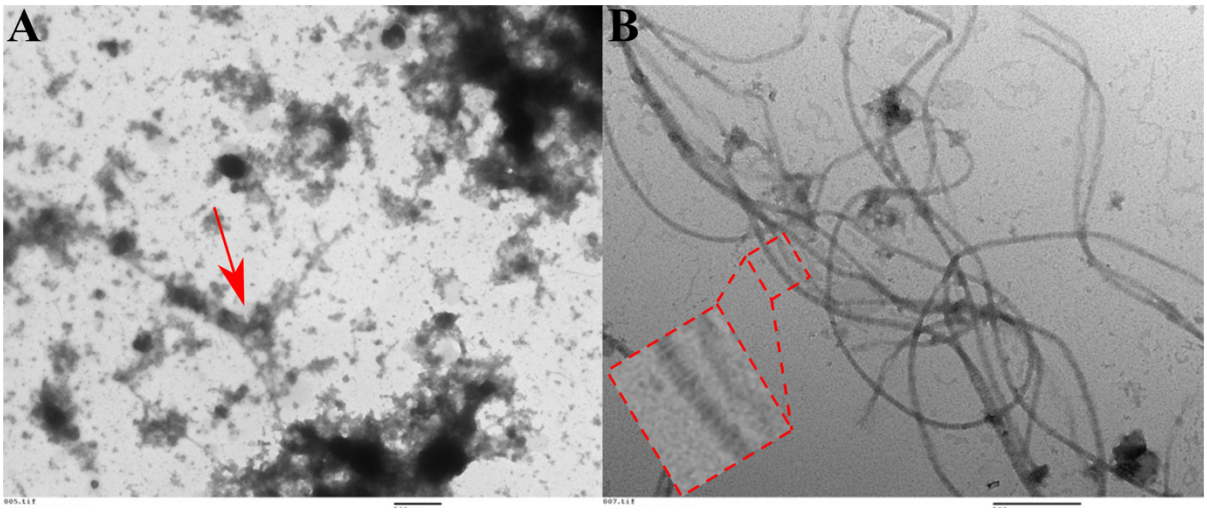


Figure 6-1 TEM images of VSMC ECM samples, negatively stained by uranyl acetate. A. Freshly decellularised VSMC ECM, showing a significant amount of NC ECM proteins. Some collagen fibrils are pointed by a red arrow, recognised by their long and straight shape. B. VSMC ECM after overnight incubation with chymotrypsin, showing collagen fibrils and their banding patterns (see enlargement on the left). Scale bars are shown underneath each image, and both of them are 500 nm long.

It is evident from TEM images that chymotrypsin is effective in digesting NC ECM proteins. Without chymotrypsin digestion, collagen fibrils are mostly covered by NC proteins and only distinguished from other ECM proteins because of their length and straightness. After chymotrypsin treatment, a significant amount of NC ECM proteins is removed from ECM

samples, and collagen fibrils can be easily identified by not only their straightness and length but also their unique banding patterns. An adequate working condition is determined to be incubating ECM with chymotrypsin overnight at 37℃ (more detailed protocol is included in section 0). However, further washing is needed to better remove proteins disassociated or loosely associated with collagen.

However, TEM has its limitations. It can be selective, because images taken can only represent the local area observed, and the information interpreted from those images may not be an appropriate representation of the whole sample. Therefore, multiple images should be taken per sample to give a more holistic understanding of the sample. Although TEM is enough for an initial assessment, a better representation and more general information are desired to study the change in chymotrypsin treated ECM samples and to make a solid conclusion about effectiveness of this treatment.

6.1.2 Fluorescence Spectroscopy

Fluorescence has long been used to study proteins because many proteins with Y, W or F residues can auto-fluoresce. It can be used to investigate a wide range of topics related to protein structure because fluorescence is sensitive and changes in response to protein structural changes, including conformational changes, altered binding activities, etc [156][157][158]. This sensitivity allows analysis of ECM component changes due to chymotrypsin treatment. More importantly, unlike TEM, fluorescence can reflect the overall features of the entire sample in favourable circumstances.

Among these three fluorescent amino acids, F is not studied as extensively as the other two because its quantum yield is too low (Table 6-2), making it unlikely to be observed when W or Y are also present. Y gives simple emission spectra. When excited at 275 nm at neutral pH, its emission maximises at 305 nm. A band at 330-350 nm sometimes appears together with its main peak, which originates from Y oxidation and is assigned as tyrosinate [159]. Tyrosinate is weakly fluorescent and can be easily lost under other peaks. W is much more often used as a fluorescence probe of protein structure than F and Y because its emission is highly sensitive to its local environments, such as polarity and solvent exposure. Commonly, when excited at 285 nm, W folded inside proteins in a non-polar environment gives a maximum emission around 330 nm; if the maximum emission wavelength is longer than 330 nm, W is in a more

polar environment, and in most cases, more exposed to surrounding solvent [160]. Because W is not present in collagen, W emission spectra are used in this project as measurements of whether chymotrypsin effectively removes W-containing proteins in ECM samples. Fluorescence experiments were developed in collaboration with Dr. Uliana Bashtanova and Ieva Goldberga.

Table 6-2 Fluorescence properties of F, Y and W in aqueous solutions at neutral pH [156][158], and wavelengths used in experiment.

Amino acid	**Quantum yield in water**	**Excitation wavelength in literature /nm**	**Excitation wavelength in experiments /nm**	**Emission wavelength in literature /nm**	**Emission wavelength in experiments /nm**
F	0.024	max at 260	-	260-340; max at 280	-
Y	0.14	max at 275	275	280-380; max at 305	285-450
W	0.13	max at 285	285	300-450; max at 350	295-450

The effectiveness of chymotrypsin treatment is measured by comparing fluorescence emission spectra of W and Y between differently treated ECM samples and pure collagen type I (from bovine Achilles tendon), in particular comparing parameters such as band shapes and wavelength shifts. Absolute fluorescence intensity cannot be used to quantify the amount of W or Y containing proteins in ECM, because measured intensity depends on many factors, the majority of which cannot be determined in this experiment. For example, a quantitative correlation between fluorescence intensity and the amount of fluorescent proteins requires at least sample homogeneity along the optical beam path through the sample; however, with ECM samples, it is difficult to guarantee the sample is homogenous across the entire sample; and therefore, fluorescence on ECM samples can only be interpreted qualitatively rather than quantitatively. Considering W is absent in collagen, fluorescence emission spectra of W - or lack of W emission - can indicate the purity of the collagenous ECM sample, in other words, the effectiveness of NC protein removal. Typical fluorescence emission spectra, both Y and W, of ECM samples before and after chymotrypsin treatment are shown in Figure 6-2, as well as emission spectra from pure collagen type I for comparison.

The same chymotrypsin digestion procedure was repeated on a different in vitro ECM harvested from foetal sheep osteoblasts (FSOBs) cell cultures to test if the procedure is generic and robust. When excited at 275 nm (to observe primarily Y) and 285 nm (to observe primarily W), emission spectra of both VSMC and FSOB ECM look very similar (all four black spectra in Figure 6-2), showing a broad emission peak with a maximum wavelength around 330 nm. This maximum implies that both ECM samples have a significant amount of folded W. The broadness of the emission peaks indicates that W distributes in a variety of different local environments in ECM. Signals from Y are presumably overlapped by this broad, intense peak from W.

After incubating ECM samples with chymotrypsin, W emission spectra, i.e. spectra arising from 285 nm excitation, (two red spectra in Figure 6-2 on the right) change dramatically from the previously broad peak centred around 330 nm to a decay curve with no obvious maximum. This change suggests that a substantial amount of W-containing proteins is removed from this sample, changing the emission peak shape. Two humps are observed around 310 nm and 325 nm, which might be from remaining W in different environments. In Y emission spectra, i.e. spectra arising from 275 nm excitation, (two red spectra in Figure 6-2 on the left), ECM samples start showing distinct Y peaks at around 300 nm after chymotrypsin treatment. Appearance of the Y peak can result from significant decreasing W-containing proteins/W emission, and thus, uncovering hidden Y signals, which are expected from collagen as well as other ECM proteins. Both effects, significant decline in W peak intensity and appearance of the Y peak, suggest successful removal of W-containing proteins and the effectiveness of chymotrypsin in digesting minor ECM proteins.

A washing step was introduced after the overnight chymotrypsin treatment to better remove remaining proteins, using a gentle detergent Tween-20 [161]. This extra washing step further smooths the emission spectra of collagenous ECM samples (all green spectra in Figure 6-2) and makes their spectra more like the emission spectra of pure collagen (blue spectra in Figure 6-2). This washing seems to be more effective to VSMC ECM than FSOB ECM: the line shape does not change as much in FSOB as in VSMC.

However, information obtained from fluorescence emission spectra may not apply to the entire sample as whether chymotrypsin digestion is effective only on ECM surface or on the entire sample is still an open question. Concerns arise from the observation of reflected signals. The

light beam used for fluorescence excitation does not penetrate through ECM samples completely, and part is reflected via the sample surface, which is the limitation of detection for solid samples. Peaks at 360nm, 375 nm and 420 nm for λ_{ex}=275 nm, and 360 nm and 390 nm for λ_{ex}=285 nm appear in all spectra obtained (Figure 6-2), and are assigned as reflection peaks due to samples' solid nature. This means (1) ECM samples are rigid enough to reflect light, (2) unlike homogeneous protein solutions, ECM samples may not experience the same beam strength equally across the sample – some parts are penetrable and some not. For areas causing reflection, the light beam only reaches sample surface and discloses structural information of the sample surface. So, it is possible that minor proteins are hidden behind the reflective areas and are undetected by fluorescence.

To sum up, fluorescence emission spectra of chymotrypsin digested and detergent washed ECM samples show that this combination successfully removes the majority of ECM proteins. However, due to the intrinsic limitation of fluorescence measurement on solid samples, a question regarding whether chymotrypsin is able to penetrate the whole ECM structure and cleave all minor proteins is still left open.

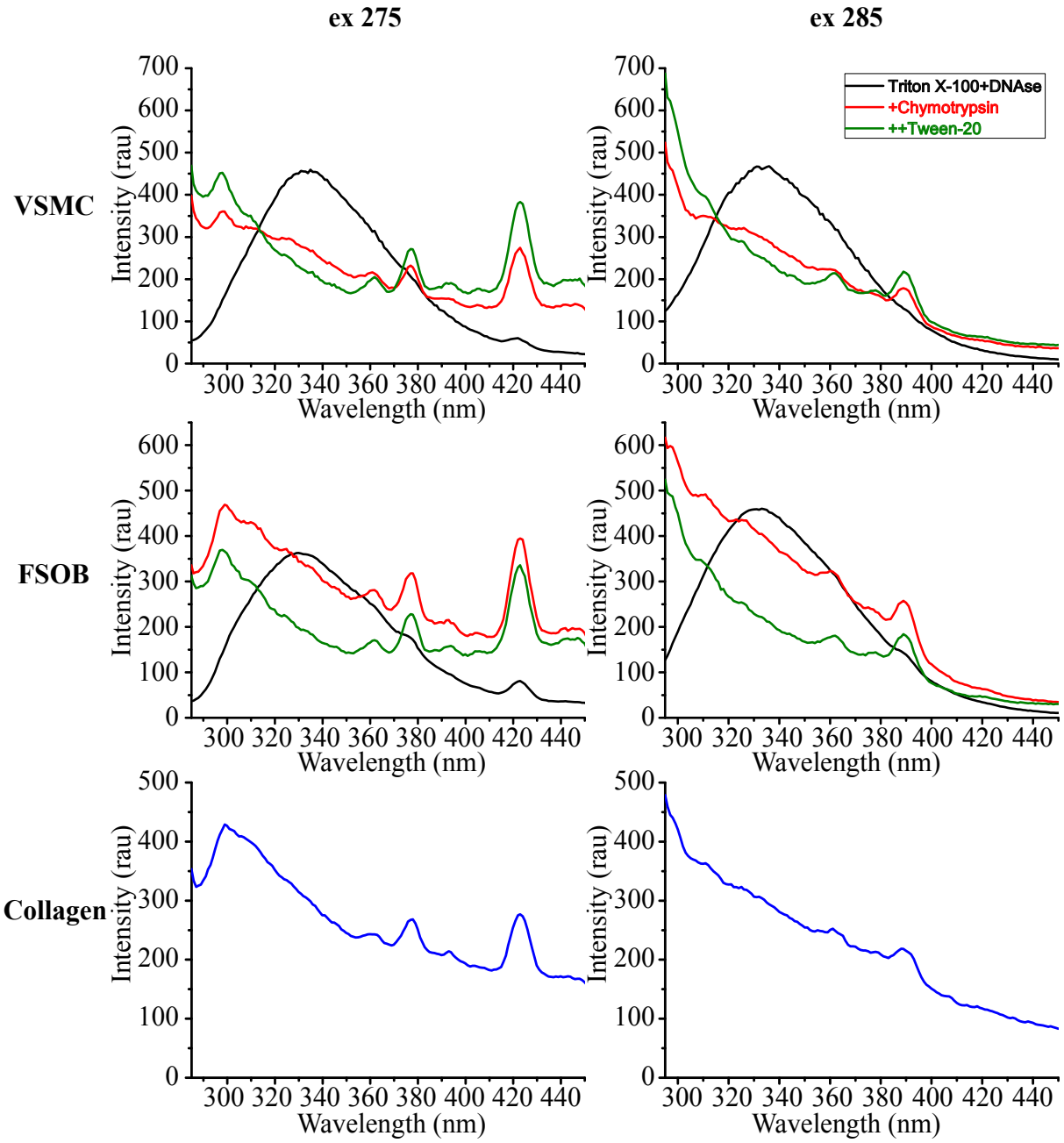


Figure 6-2 Fluorescence emission spectra acquired at different excitation wavelengths on differently treated ECM samples. Top: emission spectra from differently treated VSMC ECM at excitation of 275 nm (for Y) and 285 nm (for W). Middle: emission spectra from differently treated FSOB ECM at excitation of 275 nm and 285 nm. Spectra in black are from untreated ECM; spectra in red are from chymotrypsin digested ECM; spectra in green are from chymotrypsin digested and Tween-20 washed ECM. Bottom (blue): emission spectra of collagen at excitation of 275 nm and 285 nm.

6.1.3 Solid-State NMR (SSNMR) Spectroscopy

SSNMR provides averaged structural information of the entire sample. It is able to reveal amino acid composition of protein samples, and is also sensitive to protein secondary structures, giving different chemical shifts for residue signals in different secondary structures [162]. In particular, C_α and C' ^{13}C chemical shifts of residues in the collagen triple helix conformation are characteristically lower than those in all other secondary structures. Therefore, NMR can tell not only whether there is an overall amino acid composition change but also if the secondary structure has changed, in this work, if collagen is still present in intact triple helices.

Figure 6-3 compares 1D ^{13}C CPMAS SSNMR spectra of pure collagen type I and differently treated ECMs. ^{13}C signals from G C_α, P C_α, O C_δ, G C' and P C' become more prominent after the clean-up procedure, and overall spectrum details look more like pure collagen. This trend indicates ECM samples become more like pure collagen as a result of removing the majority of NC ECM proteins. The top three spectra in Figure 6-3, acquired on pure collagen type I, "cleaned" FSOB ECM (called "FSOB Chymotrypsin + Tween-20" in the figure, and will be referred as "Col(FSOB)" in the following text) and "cleaned" VSMC ECM (called "VSMC Chymotrypsin + Tween-20" in the figure, and will be referred as "Col(VSMC)" in the following text) look almost the same with slightly better resolution in the spectra of Col(FSOB) and Col(VSMC). However, whether there is any NC protein left is difficult to tell because SSNMR may not be sensitive enough to spot remaining minor proteins. It is also interesting to see the bottom four spectra in Figure 6-3, acquired on detergent lysed FSOB ECM (called "FSOB Detergent" in the figure, and will be referred as "ECM(FSOB)" in the following text), detergent lysed VSMC ECM (called "VSMC Detergent" in the figure, and will be referred as "ECM(VSMC)" in the following text), freeze-thaw lysed FSOB ECM (called "FSOB Freeze-thaw" in the figure, and will be referred as "ECM(FSOB) FTL" in the following text) and freeze-thaw lysed VSMC ECM (called "VSMC Freeze-thaw" in the figure, and will be referred as "ECM(VSMC) FTL" in the following text), are very similar in terms of the overall signal shapes and signal relative intensities; even though some spectra are noisier than the other due to limited sample volume or fewer number of scans (consecutive acquisitions). This similarity also indicates SSNMR is not sensitive enough to distinguish ECM decellularised from different methods.

As mentioned earlier in Table 6-1, collagen has some cleaving sites which makes it also susceptible to chymotrypsin. An ECM sample was prepared by carrying out three cycles of chymotrypsin digestion and detergent wash to test if collagen is still intact. A positive control was also made using collagenase-digested collagen to show what the ^{13}C NMR spectrum of denatured collagen looks like (Figure 6-4). ^{13}C chemical shifts and peak widths of G C_α, P C_α, O C_δ, G C' and P C' signals were inspected to see if there was any change correlating to secondary structure changes in the chymotrypsin treated sample compared to pure collagen or collagenases-digested collagen. Looking at black dotted lines in Figure 6-4, no obvious chemical shift changes are observed in the chymotrypsin-treated ECM sample (the olive green, middle spectrum in Figure 6-4), even after three cycles of chymotrypsin incubation, compared to pure collagen (the blue, top spectrum in Figure 6-4), indicating the majority of collagen triple helices are still intact. However, in the collagenase-treated collagen sample (the purple, bottom spectrum in Figure 6-4), significant signal shifts of G C_α and P C_α as well as signal broadening of all G C_α, P C_α, O C_δ, G C' and P C' signals are observed, indicating the disruption of collagen triple helix.

To sum up, the major protein left in ECM samples after the clean-up procedure is collagen. Based on the conserved ^{13}C chemical shifts and the linewidth of G C_α and C' and P C_α and C' signals, collagen is still in triple helical form. However, the sensitivity of SSNMR can be arguable; not observing a species under SSNMR does not mean absolute absence. CP SSNMR experiments can be selective in terms of detecting proteins; for instance, CP is very inefficient for mobile molecules, and so signals from such molecules are unlikely to be observed in CP spectra.

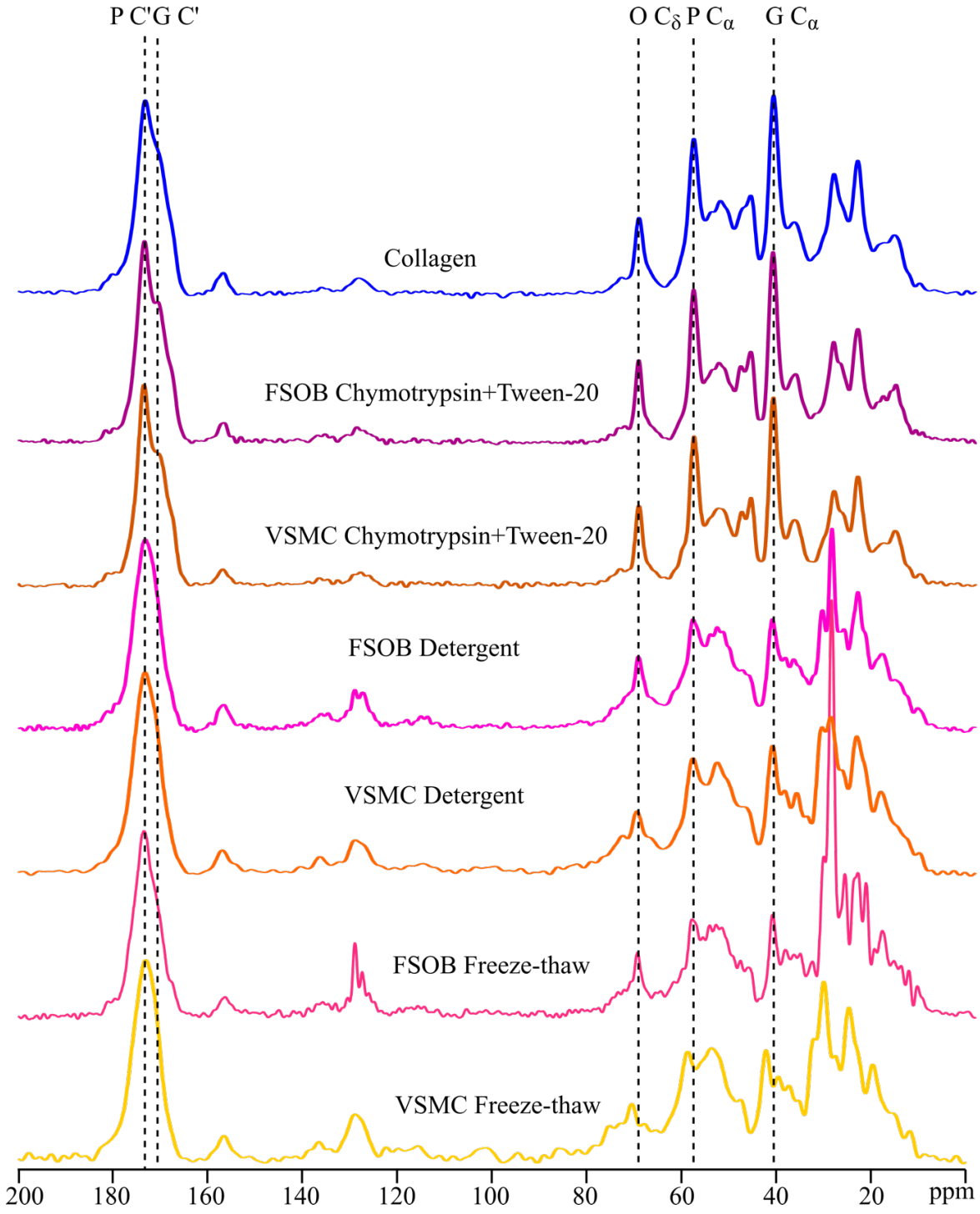


Figure 6-3 ^{13}C CPMAS SSNMR spectra of, from top to bottom, pure collagen type I (ns = 4096), Col(FSOB) (FSOB ECM chymotrypsin + Tween-20, ns = 32468), Col(VSMC) (VSMC ECM chymotrypsin + Tween-20, ns = 6144), ECM(FSOB) (FSOB ECM detergent, ns = 8192), ECM(VSMC) (VSMC ECM detergent, ns = 8192), ECM(FSOB) FTL (FSOB ECM freeze-thaw, ns = 8192) and ECM(VSMC) FTL (VSMC ECM freeze-thaw, ns = 8192). Black dash lines connect P C', G C', O C_δ, P C_α, and G C_α signals across all spectra.

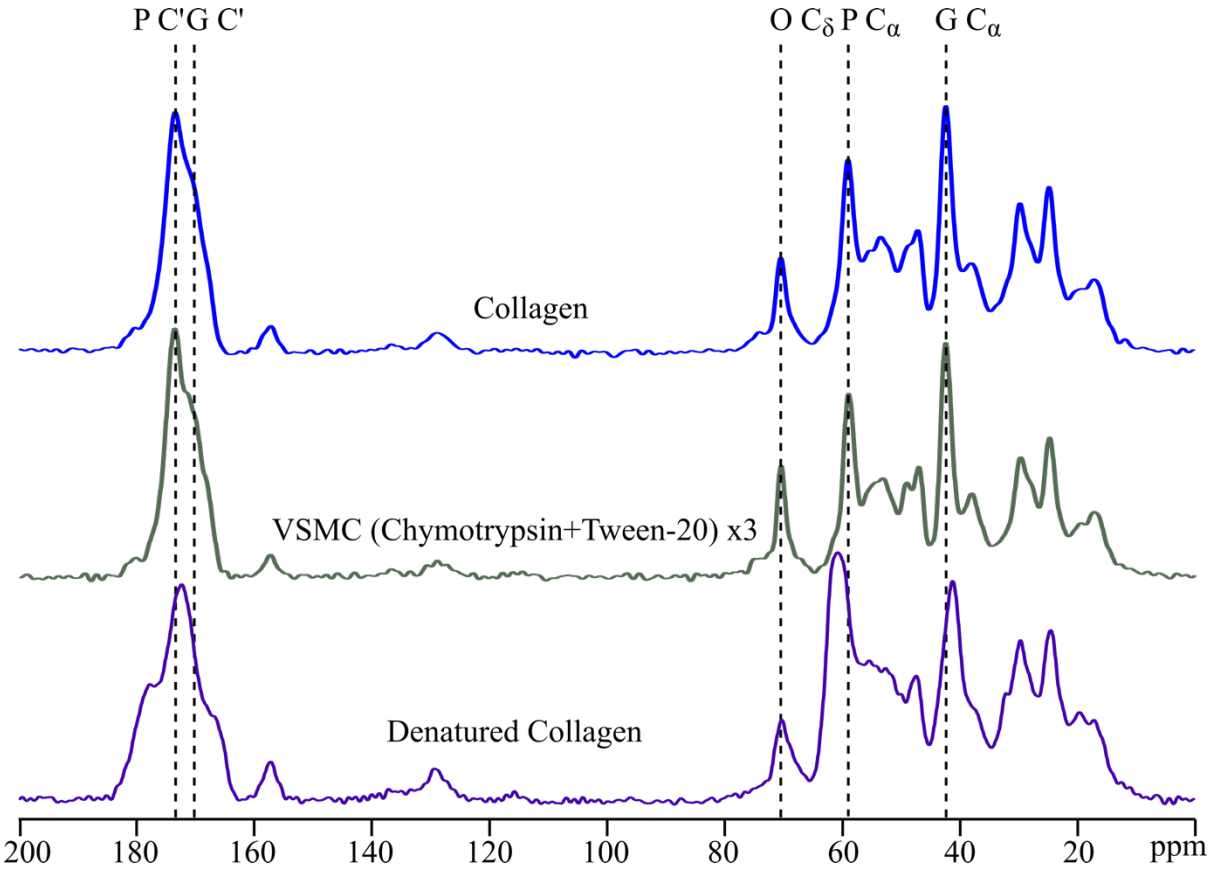


Figure 6-4 A comparison of 1D ^{13}C CPMAS SSNMR between pure collagen type I (top, blue, ns = 4096), detergent lysed VSMC ECM with 3 cycles of chymotrypsin digestion and detergent wash (middle, olive green, ns = 8192) and collagenase denatured collagen type I (bottom, purple, ns = 10240). Denatured collagen shows observable signal shifting and the sample was prepared by Dr. Sneha Bansode.

6.1.4 Amino Acid Analysis (AAA)

Knowing G is every third residue in collagen triple helix, the molar ratio of G in collagen should be around 33%. This number can vary slightly in different types of collagen and collagen from different animal species. A second special feature of collagen is the presence of O, a unique post-translational modification (PTM) in collagen (and elastin). Considering the ECM as a whole, where G and O are rarer, the presence of other proteins dilutes the molar ratio of G and O. This means removing NC proteins in ECMs should result in increased molar ratio of G and O and decreased molar ratio of L, F and Y, which are relatively less abundant in collagen compared to other ECM proteins, towards their molar ratios in collagen. AAA is able to measure the molar ratio of individual amino acids in ECM samples and provides a more quantitative perspective on amino acid composition in collagenous ECM samples, compared to the above-mentioned techniques. AAA is provided as a service by the Department of Biochemistry, University of Cambridge; raw data were sent to us in the form of molar ratio of each amino acid and were further processed for the purpose of the work by Ieva Goldberga and me.

Results from AAA (partial) are present in Table 6-3 and Figure 6-5. Pure collagen type I has almost 35% G, similar to the theoretical value of 33%, and 11% O. In terms of FSOB ECM, each step of treatment, except for the last washing step, results in growing G% and O% from 14.6%, 23.3% to 31.8% and 2.4%, 6.3% to 9.8% separately and declining L%, F% and Y% from 7.4%, 5.0% to 3.1%, 3.2%, 2.2% to 1.6% and 2.5%, 1.4% to 0.6% respectively. Reconciling these results with fluorescence data (Figure 6-2in section 6.1.2), this final detergent washing does not seem to make a difference to FSOB ECM composition. For VSMC ECM, the G% and O% barely changes with different decellularisation methods, freeze-thaw or detergent lysis, both around 18% and 3.8%; while, VSMC ECM respond well to the final detergent wash, giving final G and O molar ratios of 32.8% and 9.8%.

After all treatments, both FSOB and VSMC ECM samples are very similar to each other and to pure collagen type I, regarding G, O, L, F and Y molar ratios. Small differences are observed between ECM samples and pure collagen type I. This difference can be from a number of factors, in addition to the possibility of incomplete removal of ECM NC proteins. For example, ECM contains different types of collagen, and therefore, depending on the composition of

collagen types, amino acid molar ratios can be different. The pure collagen is derived from tendon, expected to be very largely collagen type I; that from osteoblast ECM, similarly mainly collagen type I; while that from muscle cell ECM, expected to contain significant amounts of collagen type III in addition to collagen type I. Also, tendon and muscle cell ECM are from bovine whilst the osteoblast ECM here is from sheep. Therefore, small differences in amino acid compositions between collagen in the three samples compared above are expected. It is worth mentioning here that VSMC ECM suffers more significant sample volume loss than FSOB ECM due to the clean-up procedure, indicating VSMC ECM has more NC proteins or FSOB ECM is less penetrable to the clean-up process.

Overall, the proposed clean-up procedure using chymotrypsin digestion and Tween-20 washing is effective in removing NC ECM proteins and extracting pure collagen from in vitro ECMs. Considering the sample volume involved in each measurement, usually less than 1 mg per measurement, assessment of the "cleaned" ECM samples can only represent a small portion of the total material obtained after the clean-up process. Therefore, when performing AAA, more repeats per sample are recommended to make an average and standard deviation to assess if differences in molar ratios between samples are significant or negligible, especially if the extraction product is not homogenous.

Table 6-3 Molar ratio of G, O, L, F, Y from pure collagen type I and differently-treated ECM samples, measured by AAA. Theoretical values are calculated from bovine collagen type I sequence published on UniProt [153]. [1]This O% is deducted from chicken collagen type I and can be different from bovine collagen type I.

Sample Type		G%	O%	L%	F%	Y%
Theoretical Value		33.05	6.06[1]	2.48	1.33	0.41
Collagen		**34.97**	**11.16**	**2.74**	**1.45**	**0.44**
FSOB	**Freeze-thaw**	**14.57**	**2.41**	**7.36**	**3.15**	**2.48**
	Detergent	**23.33**	**6.31**	**5.02**	**2.23**	**1.44**
	Chymotrypsin	**31.83**	**9.79**	**3.05**	**1.60**	**0.60**
	Chymotrypsin + Tween-20	**31.71**	**9.15**	**3.08**	**1.60**	**0.61**
VSMC	**Freeze-thaw**	**17.91**	**3.85**	**6.38**	**2.91**	**2.19**
	Detergent	**17.95**	**3.80**	**6.42**	**2.60**	**2.10**
	Chymotrypsin	**31.61**	**9.64**	**3.25**	**1.56**	**0.66**
	Chymotrypsin + Tween-20	**32.75**	**9.75**	**2.76**	**1.42**	**0.39**

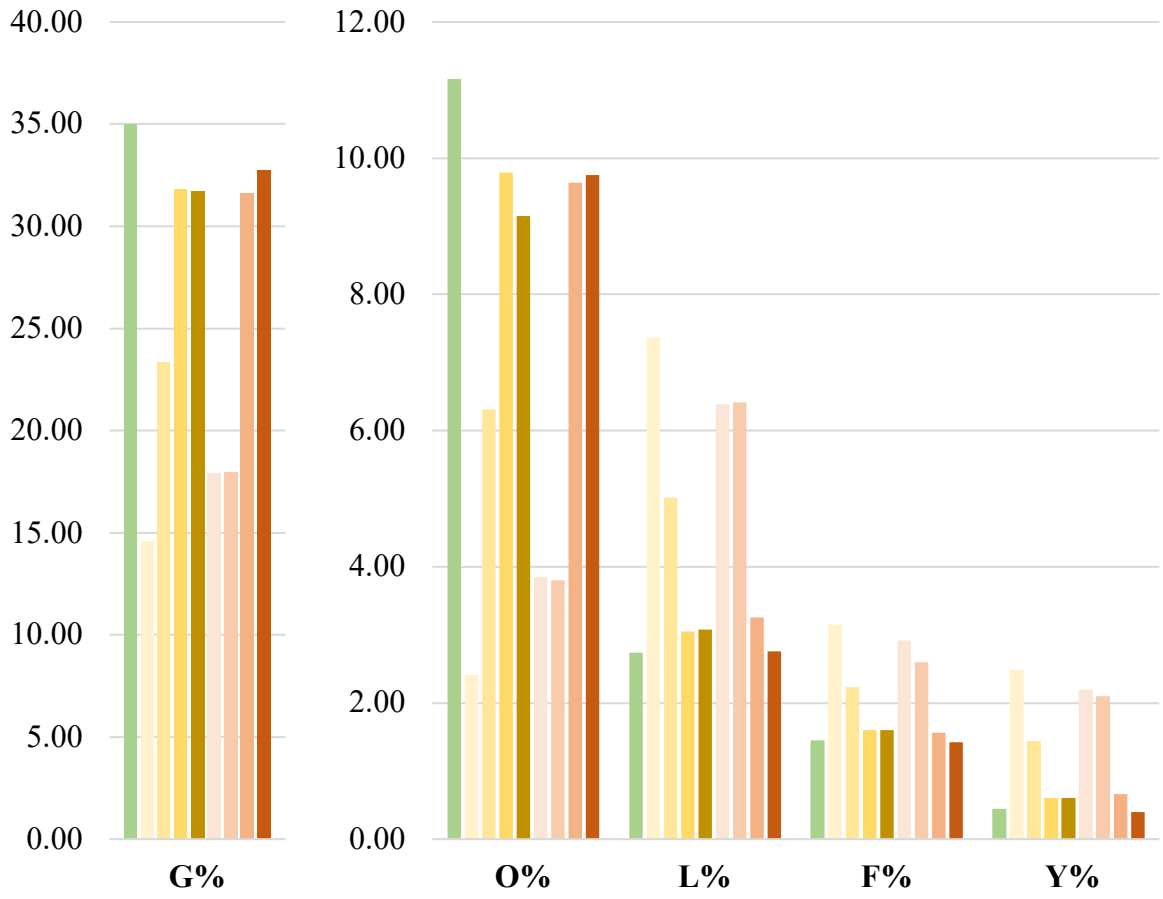


Figure 6-5 Molar ratio of G, O, L, F, Y from pure collagen type I and differently treated ECM samples, measured by AAA. Colour code is consistent with Table 6-3.

6.2 Summary

ECM is known to have many proteins in minor quantities other than collagen. Extracting collagen from integral in vitro ECM can be complicated because ECM has a complex organisation and common proteolytic enzymes may denature collagen at the same time as digesting minor ECM proteins. Here, a protocol using a combination of detergent cell lysis, chymotrypsin digestion and detergent washing is proposed, and successful extraction of pure and intact collagen from complex in vitro ECM is achieved. The effectiveness of the clean-up protocol was examined by four analytical techniques, including TEM, fluorescence spectroscopy, SSNMR spectroscopy and AAA.

TEM images offer a direct and visual examination of chymotrypsin treated ECMs. It is clear from the TEM images that as a result of the clean-up procedure, less NC proteins are present in ECM samples, and less proteins are bound to collagen fibrils, and collagen fibrils are more clearly defined as a result. Additionally, TEM images can be used to assess whether collagen fibrils left are intact. However, TEM can be selective because TEM images only cover a very limited area, relative to a whole TEM grid and a tiny fraction of the sample, relative to the total sample. This limitation can be overcome by (1) preparing more TEM grids per sample (2) taking images across the whole grid and (3) analysing as many images as possible before making a conclusion.

Fluorescence emission spectra were acquired on ECM samples with and without chymotrypsin treatment and compared with spectra acquired on pure collagen type I were also used to assess the effectiveness of the clean-up protocol on eliminating ECM proteins. It is clear that after the clean-up protocol, the fluorescence emission spectra of ECM samples become more like those of pure collagen, suggesting the success of extracting collagen and removing NC proteins. However, it is uncertain whether ECM samples are homogenous across a measuring cuvette along the beam pathway in the fluorimeter. Therefore, it is possible that information obtained on solid ECM samples only reflects a specific portion of the sample, such as on the sample surface, rather than the whole sample.

Of all techniques used, SSNMR spectroscopy is the only one able to disclose structural and compositional information of an entire ECM sample. Though the sensitivity of SSNMR

spectroscopy is lower than other techniques, it is not difficult to conclude from the comparison of SSNMR spectra obtained from differently treated samples that chymotrypsin successfully removes ECM proteins and the majority protein left after the series of treatments is intact triple helical collagen.

AAA is the most quantitative analysis involved compared to the other three. It reveals information on amino acid compositions of ECM samples, showing that molar ratios of certain amino acids in cleaned-up ECM become closer to those of pure collagen. Its results can also be selective due to the small volume used per measurement. Therefore, information obtained may not be truly reflective of the whole sample. This limitation can be controlled by measuring multiple portions from one sample and taking an average value and examining the standard deviation to understand the degree of compositional heterogeneity in the sample.

Though each technique has its limitations, a combination of all techniques assembles a full picture of the ECM sample after the clean-up steps. Therefore, a final conclusion is made that the designated clean-up procedure is effective in extracting collagen from in vitro ECM. This extraction method will be used to generate isotope enriched collagen, and study specific amino acid behaviour in glycation reactions.

It is noticed from this work that (1) as assessed by the AAA results, VSMC ECM composition is not affected by decellularisation methods, either freeze-thaw lysis or detergent lysis, as much as FSOB ECM, giving ECMs that are slightly different composition after detergent lysis. AAA offers a more sensitive and different perspective from SSNMR spectra in this regard. (2) The final detergent washing step does not clean FSOB ECM as much as it does VSMC ECM, which is seen from the analysis of fluorescence spectra and AAA, and thus, may not be necessary for FSOB ECM. (3) More significant sample volume loss was observed on VSMC ECM than FSOB ECM, indicating VSMC ECM has smaller initial collagen content or is more penetrable by chymotrypsin. These different trends between two different types of ECM might be a reflection of their intrinsic differences in architecture and composition because they are from two different cell types.

7 Identification of Major Glycation Products in Model Systems

The glycation reaction has been studied for almost a century because of its association with different pathologies. Many glycation products have been identified in in vivo tissues and different in vitro model systems, ranging from protected amino acids to collagen. However, a bridge between in vitro models and in vivo tissues is lacking to justify the biological relevance of glycation products detected in in vitro models. In this project, in vitro extracellular matrix (ECM) is used as the bridging in vitro model to mimic native tissues. The resemblance between the in vitro ECM models used here and native tissues has already been established [120].

A number of characterisation techniques, including fluorescence spectroscopy, mass spectroscopy coupled with high performance liquid chromatography or gas chromatography and immunohistochemical staining, have been deployed to study the glycation chemistry and to identify glycation products in model systems. However, those techniques intensively used so far have their intrinsic disadvantages and potentially introduce artefacts into the study and reveal information not representative to the physiological conditions. SSNMR spectroscopy can overcome some of those disadvantages as it only requires minimum sample preparation so largely preserves biological samples in their native states. Therefore, SSNMR spectroscopy is used as the primary characterisation technique in this project. In order to overcome the sensitivity problem experienced by SSNMR, isotope enrichment is heavily used in this work.

Ribose-5-phosphate is used as the main glycating agent in this project because firstly its glycation rate is faster than glucose and ribose and therefore able to generate more AGEs over the same period of time, and secondly it is an important metabolite in nucleic acid metabolism. Glucose and ribose are also used in this project as benchmarks for comparing glycation rates and glycation products with ribose-5-phosphate.

7.1 Glycation Products Derived from Pure Collagen Type I with U-$^{13}C_6$-Glucose and U-$^{13}C_5$-Ribose-5-Phosphate

Collagen type I, as the major component of the ECM, plays a crucial role in ECM structure integrity and functions (reviewed in section 2.1). Pure collagen type I has been used as a simplified model of native tissue ECM to study the effects of glycation and used in this project as a starting point for further exploration of more complicated glycation models.

This section will discuss possible glycation products derived from pure collagen type I (from bovine Achilles tendon) with U-$^{13}C_6$-glucose (U-$^{13}C_6$-Glc) and U-$^{13}C_5$-ribose-5-phosphate (U-$^{13}C_5$-R5P). Collagen glycation using U-$^{13}C_5$-ribose (U-$^{13}C_5$-Rib) has been studied by other colleagues before [163][136], and some results have been published [163]; therefore, the experiment will not be repeated here.

7.1.1 U-$^{13}C_6$-Glc

Glucose is the most bio-relevant glycating agent because it is circulating in human bodies all the time. The concentration of circulating glucose can fluctuate in non-diabetic human bodies between 3.9-5.6 mM, and to above 10 mM in hyperglycaemia [164]. It is known to be a slow glycating agent [121][165] as the glycation rate is proportional to the relative concentration of glucose open chain aldehyde (0.004% [166]) form versus cyclic form in solution (Figure 7-1).

Pure collagen type I was incubated with 100 mM U-$^{13}C_6$-Glc for four months (details are included in section 5.3) SSNMR analysis. It is clear from Figure 7-2 that new species are generated over the incubation and give ^{13}C NMR signals at 180 ppm, 165 ppm, 96 ppm, 93 ppm, 75 ppm, 70 ppm and 63 ppm (see the black dotted lines in Figure 7-2). In addition, glycation products are generated, giving a broad signal in the amide region around 170 ppm (shown in the purple spectrum in Figure 7-2). Signals between 60-100 ppm are likely to be from hydroxylated carbons in sugar adducts, in other words, early glycation products. More specifically, signals at 96 ppm and 93 ppm share similar chemical shifts as glucose C1 in β- and α- pyranose conformations (structures **b** and **a** in Figure 7-1). These two signals are likely to be from unreacted glucose remaining in the sample or possibly from carbon atoms in very similar structures. The broadness of all the sugar-like signals observed within the 60-100 ppm

region possibly indicates that glucose adducts are in multiple conformations and environments. The observation of sugar-like glycation intermediates after four months incubation reflects the slow reaction rate of glucose as a glycating agent.

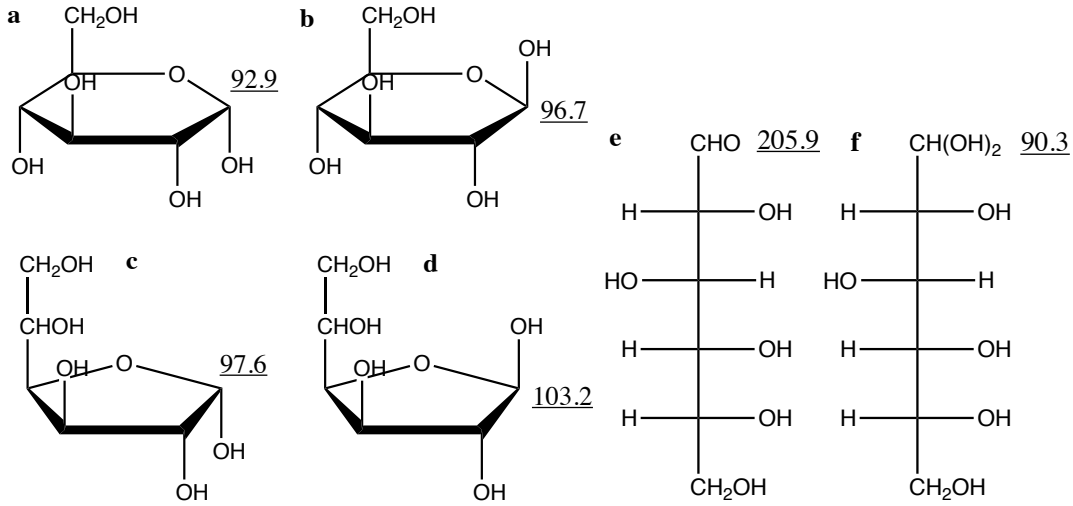


Figure 7-1 Isomers of glucose detected in solution-state NMR, their relative concentrations and C1 ^{13}C chemical shifts. a. α-pyranose (37.64%). b. β-pyranose (61.96%). c. α-furanose (0.11%). d. β-furanose (0.28%). e. open chain aldehyde (0.004%) and f. its hydrate form (0.0059%). [166]

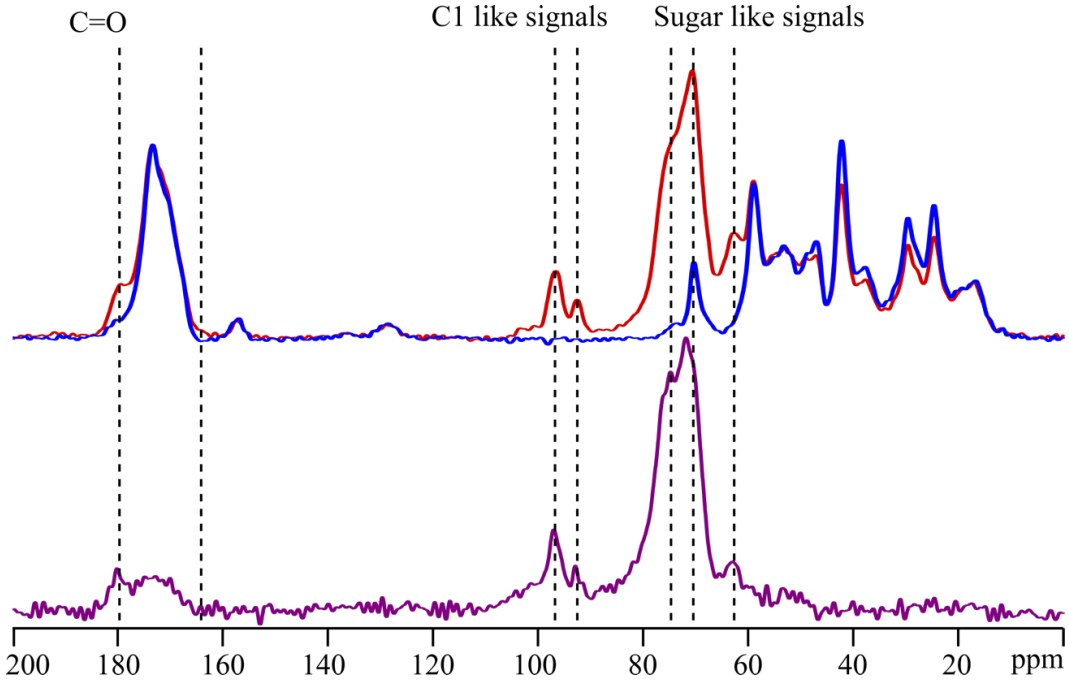


Figure 7-2 Stacking of 1D ^{13}C CPMAS SSNMR spectra of pure collagen type I (blue, ns = 4096) and U-$^{13}C_6$-Glc glycated collagen type I (red, ns = 5472) and 1D ^{13}C PC71D SSNMR spectrum, only revealing ^{13}C-enriched species, of U-$^{13}C_6$-Glc glycated collagen (bottom, purple, ns = 10240). Dotted line connects glycation products. This glycated sample was prepared by Miss Holly Chetwood.

The observation of sugar-like glycation intermediates as the dominant glucose glycation products suggests that other signals might also come from early stage glycation products rather than AGEs. Thus, we propose that the ^{13}C signal at 165 ppm might be from carbon-13 nuclei in Schiff base like structures (>^{13}C=N-), and the signal at 180 ppm might be from the ketone functional group in Amadori products or other intermediates with aldehyde functional groups. These two signals indicate that glycation in this model system might still be at the reversible stage, but the generation of ketone/aldehyde functional groups should lead to further glycation reactions and eventually form AGEs.

In order to assign the ^{13}C signals from U-$^{13}C_6$-Glc derived glycation products more definitively, a 2D ^{13}C-^{13}C PDSD experiment was carried out. 2D ^{13}C-^{13}C PDSD experiments can detect ^{13}C-^{13}C correlation pairs close in space, and thus ^{13}C connectivity and ultimate molecular structure can be deducted from 2D ^{13}C-^{13}C PDSD spectra. Apart from ^{13}C-^{13}C correlation signals seen in the so-called "sugar region" of the ^{13}C NMR spectrum (60-100 ppm), four ^{13}C-^{13}C correlation signals are detected in the 2D ^{13}C-^{13}C PDSD spectrum (Figure 7-3), ^{13}C (45 ppm) -^{13}C (173 ppm), ^{13}C (50 ppm) -^{13}C (170 ppm), ^{13}C (71 ppm) -^{13}C (180 ppm) and ^{13}C (74 ppm) -^{13}C (178 ppm) (chemical shifts are rounded to integers and included in Table 7-1). Based on their F1 chemical shifts, signals 1 (173 ppm) and 2 (170 ppm) in the 2D ^{13}C-^{13}C correlation spectrum (Figure 7-3 and Table 7-1) are probably carbonyl carbons in carboxylate groups (-$^{13}COO^-$-) or amide groups (-^{13}CONH-), which are not likely to react further, and thus potentially from AGEs; while signals 3 and 4 (both ~ 180 ppm) could be carbonyl carbons in aldehyde groups (-^{13}COH) or ketone groups (>^{13}CO), which are still reactive and likely to be from glycation intermediates. However, without observing further connectivities to other ^{13}C in the 2D ^{13}C-^{13}C PDSD spectrum, it is difficult to conclude what exact products they are and their chemical structures.

Overall, it does not seem that AGEs are generated in a significant amount in the U-$^{13}C_6$-Glc glycated collagen sample during this incubation time, and the amount of glycation products generated is not sufficient to be observed and identified by SSNMR in this model system. Although AGEs crosslinks have been detected in collagen and glucose incubation systems by more sensitive techniques such as mass spectroscopy and fluorescence spectroscopy within a similar time scale, they are in extremely low concentration, for example, 1 pentosidine molecule per 200 collagen molecules [167] and this low concentration is below the detection limit of SSNMR.

To facilitate the glycation reaction and accumulate enough glycation products to characterise more fully by SSNMR within a reasonable time frame, U-$^{13}C_5$-Rib and U-$^{13}C_5$-R5P are used henceforth as the primary glycating agents in this project, unless otherwise specified.

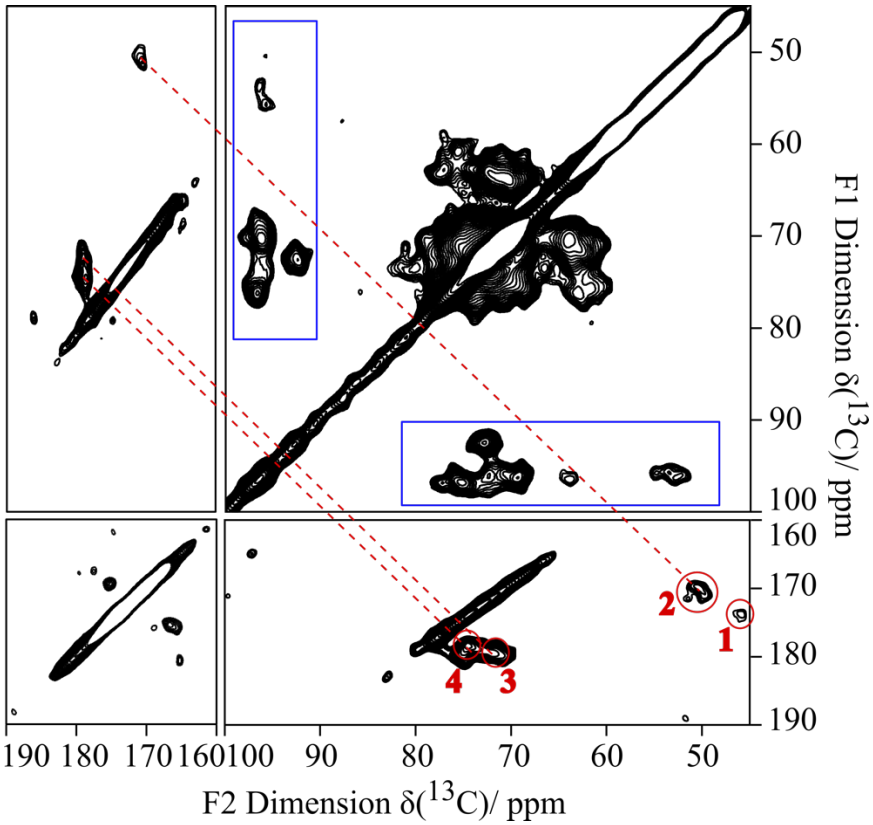


Figure 7-3 2D ^{13}C-^{13}C PDSD spectrum of U-$^{13}C_6$-Glc glycated collagen type I (20 ms mixing time, ns = 720, 218 F1 slices), showing a group of overlapping signals in the sugar region and four signals in the bottom right region. Signals included in blue rectangles are the Glc C1 signals. All correlation signals included in the right top panel fall in the "sugar region".

Table 7-1 ^{13}C-^{13}C cross peaks detected by the 2D PDSD experiment on U-$^{13}C_6$-Glc glycated collagen type I sample (indicated by red dotted lines and numerals in Figure 7-3), excluding correlation signals in the sugar region.

Signal Number in Figure 7-3	**Cross Peak Correlations**	
	Signal 1/ ppm	**Signal 2/ ppm**
1	45	173
2	50	170
3	71	179
4	74	178

7.1.2 U-$^{13}C_5$-R5P

Ribose-5-phosphate (R5P, structure shown in Figure 7-4) is an essential intermediate in nucleic acid synthesis and energy metabolism and a vigorous endogenic glycating agent. R5P generates over 5 times more fluorescent glycation products than Glc at the same concentration and within the same incubation period [165]. Apart from its reactivity towards proteins, R5P is susceptible to slow degradation at physiological temperature and pH, potentially generating many other reactive species, better mimicking the pool of various glycating agents in vivo. Figure 7-5 shows 1D ^{13}C solution state NMR spectra of 6-week incubated U-$^{13}C_5$-R5P and U-$^{13}C_6$-Glc in phosphate buffer respectively. Much more intense signals in the carbonyl region are observed in the U-$^{13}C_5$-R5P incubation. Those signals are attributed to R5P in its open chain form and R5P degradation products, indicating that R5P can glycate faster than Glc and potentially generate other reactive glycating agents.

Pure collagen type I was incubated with 50 mM U-$^{13}C_5$-R5P for two months. Even though both concentration and incubation time are halved compared to what was used in making the previous U-$^{13}C_6$-Glc glycated collagen sample, a bigger variety of glycation products are generated after R5P incubation than Glc incubation (Figure 7-6). In addition to ^{13}C signals already seen in the U-$^{13}C_6$-Glc glycated collagen, new signals are observed at around 190 ppm. This signal has previously been observed in U-$^{13}C_5$-Rib glycated collagen [163][136] and assigned as norpronyl-lysine. Also, a number of ^{13}C signals are observed below 60 ppm, which are likely to be from sp3 carbons, such as carbon atoms in methyl or methylene groups.

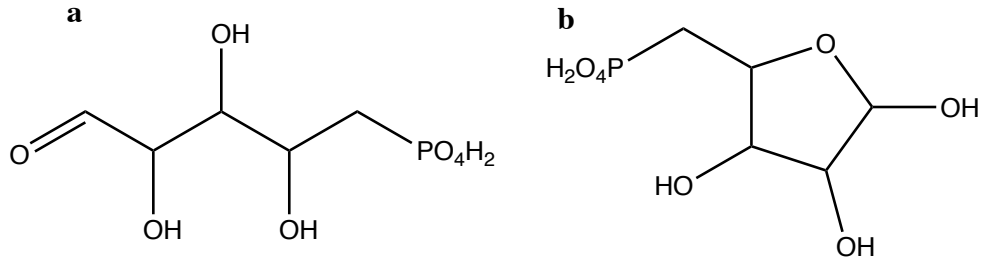


Figure 7-4 Isomers of ribose-5-phosphate. a. open chain form. b. furanose form. It can adopt either α- or β-furanose conformation. There is an equilibrium between the two forms in solution.

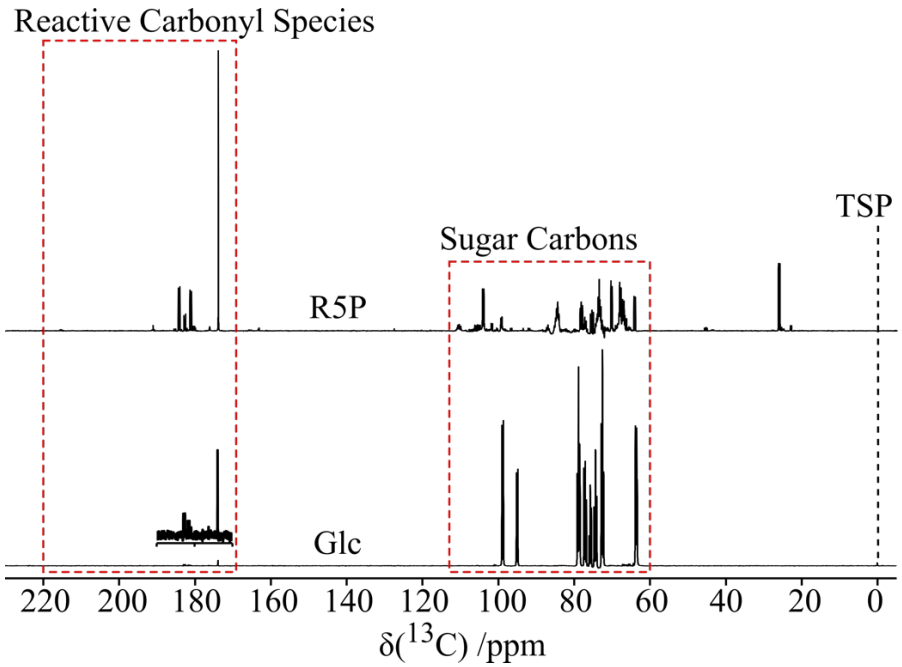


Figure 7-5 ^{13}C solution-state NMR spectra of 6 week-incubated 50 mM U-$^{13}C_5$-R5P (top) and U-$^{13}C_6$-Glc (bottom) in phosphate buffer alone. Both spectra are referenced against TSP with respect to both chemical shift and intensity.

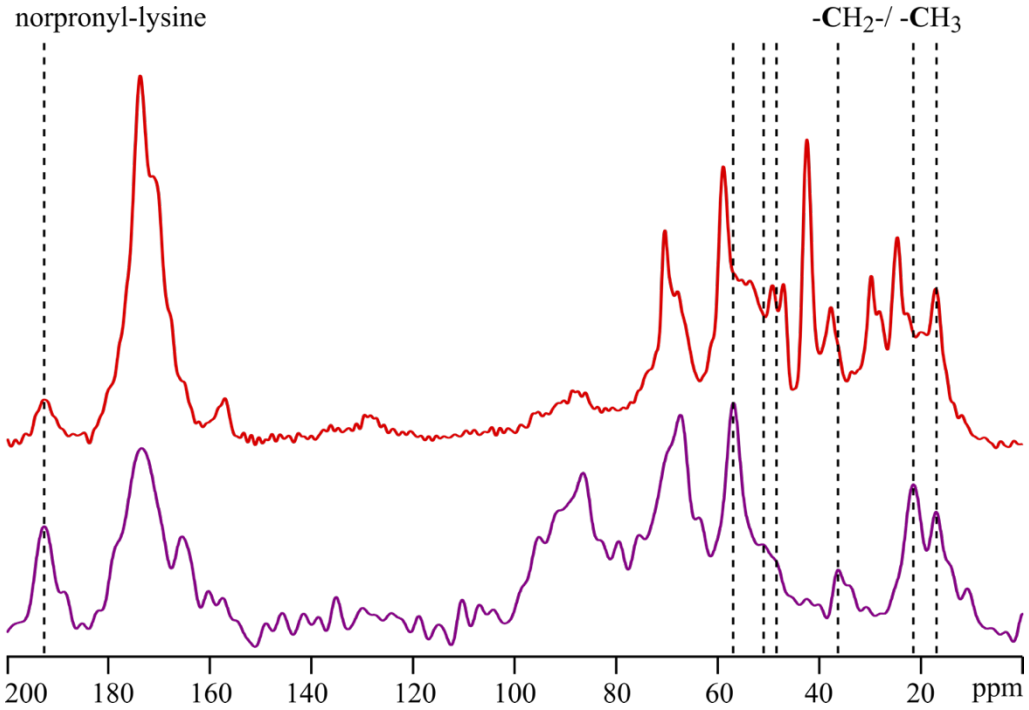


Figure 7-6 1D ^{13}C CP-MAS (top, red, ns = 1024) and PC71D (bottom, purple, ns = 2048) SSNMR spectrum of U-$^{13}C_5$-R5P glycated pure collagen type I, showing more glycation products than U-$^{13}C_6$-Glc glycated collagen presented in previous section (Figure 7-2). Dotted lines indicate glycation products generated in this incubation period.

The phosphate group on R5P is considered a good leaving group, so R5P is not expected to retain its phosphate group in the glycation products. 1D ^{31}P CPMAS SSNMR experiments were performed on the U-$^{13}C_5$-R5P glycated sample to measure phosphorus content in this sample to confirm the reaction behaviour of the phosphate group on R5P during the glycation reaction. Figure 7-7 shows that little ^{31}P content is detected after glycation. ^{31}P signals ranging between 0-5 ppm are more likely to be from inorganic phosphate left from the phosphate buffer solution used in the incubation than sugar-bonded organic phosphate monoester. Therefore, the majority of the phosphate groups originally in R5P are not present in R5P derived glycation products. When R5P reacts with proteins, the fructosellysine-equivalent Amadori product called 5-phosphono-ribuloselysine (5-PRL) forms. The negatively-charged 5-PRL might then affect the extent of glycation because of the potential repulsion with negatively-charged glycating agents, R5P. However, 5-PRL can go through further elimination to remove the terminal phosphate group and make the terminal carbon a methyl group (mechanism proposed in [163]), which generates a charge-neutral diketone intermediate. If the phosphate group is hydrolysed at any stage, R5P becomes a Rib as the "effective glycating agent", and therefore, R5P probably generates a very similar set of glycation products compared to Rib.

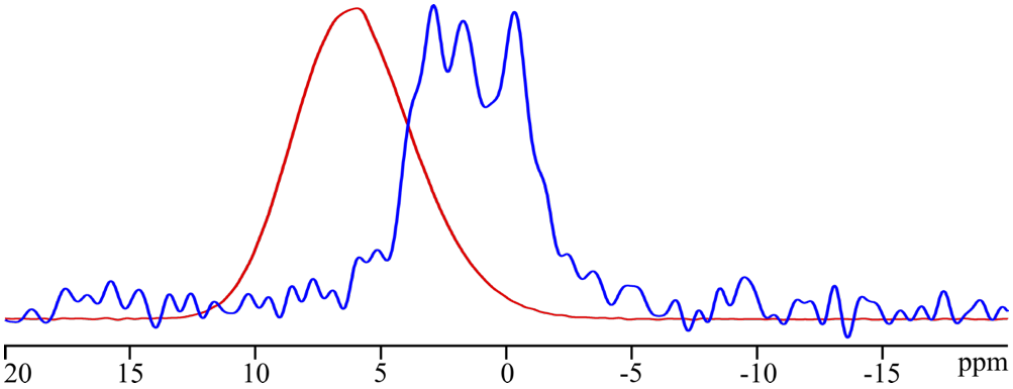


Figure 7-7 1D ^{31}P CPMAS SSNMR spectrum of R5P (red, ns = 5384) and U-$^{13}C_5$-R5P glycated collagen type I (blue, ns = 800). "ns" stands for "number of scans", meaning the number of consecutive scans acquired on the samples. More "ns" accumulates more signal intensity and, in general, can give spectra with better signal to noise ratio.

Two different types of 2D ^{13}C-^{13}C correlation experiments were performed on this U-$^{13}C_5$-R5P glycated collagen sample to gain more detailed information on the glycation products generated over incubation. All signals observed in 2D ^{13}C-^{13}C spectra are from species derived from U-$^{13}C_5$-R5P with signals from collagen invisible as 2D ^{13}C-^{13}C experiments require carbon-13 and carbon-13 interactions which are rare in non-isotope labelled collagen (natural abundance

of ^{13}C is around 1.1%). 2D ^{13}C-^{13}C PC7 experiments are good at detecting directly bonded carbon-13 pairs and revealing structural information on carbon-13 connectivity. 2D ^{13}C-^{13}C PDSD experiments detect ^{13}C-^{13}C correlations via the intervening ^{1}H network in the sample, not necessarily directly bonded, so they can detect correlated ^{13}C pairs with different distances apart, by varying an experimental parameter – the so-called mixing time (NMR terms are explained in section 4.6.1). Longer mixing time allows detection of correlated ^{13}C pairs with longer distances apart. (up to a limit; more details in chapter 4) Therefore, a combination of these two different experiments should be able to elucidate more structural information than using any of them alone.

Looking at the 2D PC7 spectrum first (Figure 7-8), it is clear to observe ^{13}C-^{13}C correlation signal pairs at, from top to bottom, 17-57 ppm, 37-59 ppm, 22-91 ppm, 21-173 ppm, 50-172 ppm, 57-176 ppm, 57-193 ppm and 165-165 ppm. The two poorly-resolved signals circled out in the centre of the spectrum are likely to be from ^{13}C-^{13}C correlations within sugar-like glycation intermediates, such as between hydroxylated carbons and hemiacetal carbons. ^{13}C-^{13}C correlation signals observed in this 2D PC7 spectrum are compared with those observed for U-$^{13}C_5$-Rib glycated collagen, and most signals are found in both spectra (Table 7-2).

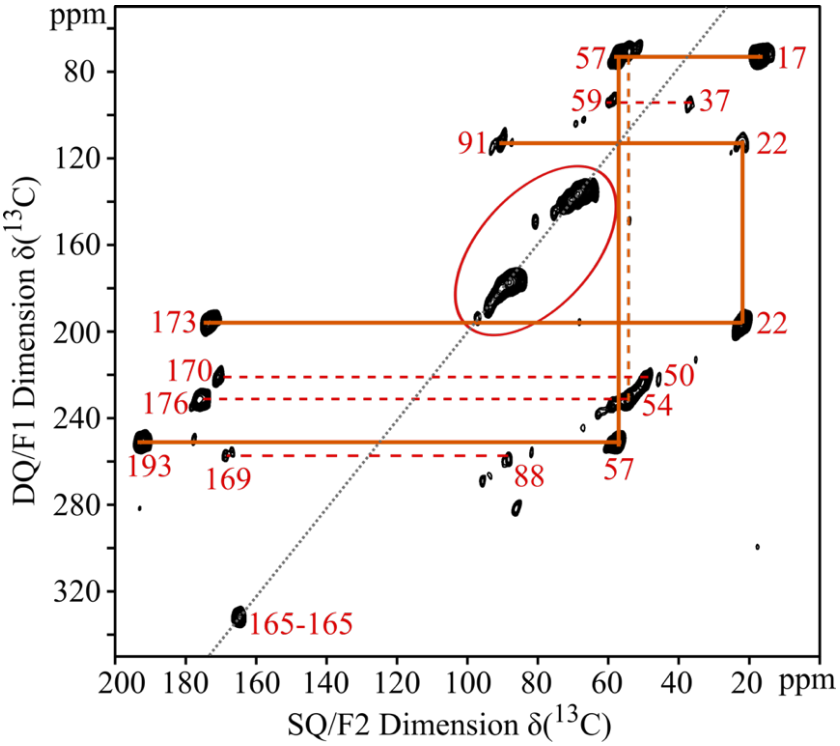


Figure 7-8 2D ^{13}C-^{13}C PC7 SSNMR spectrum of U-$^{13}C_5$-R5P glycated collagen (m = 30 mg, ns = 1024 and 70 F1 slices). Correlated signals are linked, and chemical shifts are indicated. The DQ diagonal is shown as a grey dotted line.

To gain more structural information and confirm assignments, 2D ^{13}C-^{13}C PDSD experiments with two different mixing times were performed on this sample. In theory, a variation in mixing time offers an opportunity to distinguish the spatial proximity between two correlated ^{13}C. For example, if the intensity of the cross peak is stronger in the spectrum acquired with longer mixing time than in the spectrum acquired with shorter mixing time; or in some cases, if some cross peaks are only observed in spectra acquired with longer mixing times, this pair of correlated carbon-13 giving rise to the cross peak is relatively far apart and needs a longer time to accumulate signal intensity. However, in this case, the PDSD spectra (Figure 7-9) acquired with shorter, 20 ms, and longer, 100 ms, mixing time are not very different. Most signals seen in the 100 ms spectrum (the blue spectrum in Figure 7-9) are also observed in 20 ms spectrum (the red spectrum in Figure 7-9), except for 24-180 ppm, 69-193 ppm and 158-166 ppm (indicated by black dash lines in Figure 7-9).

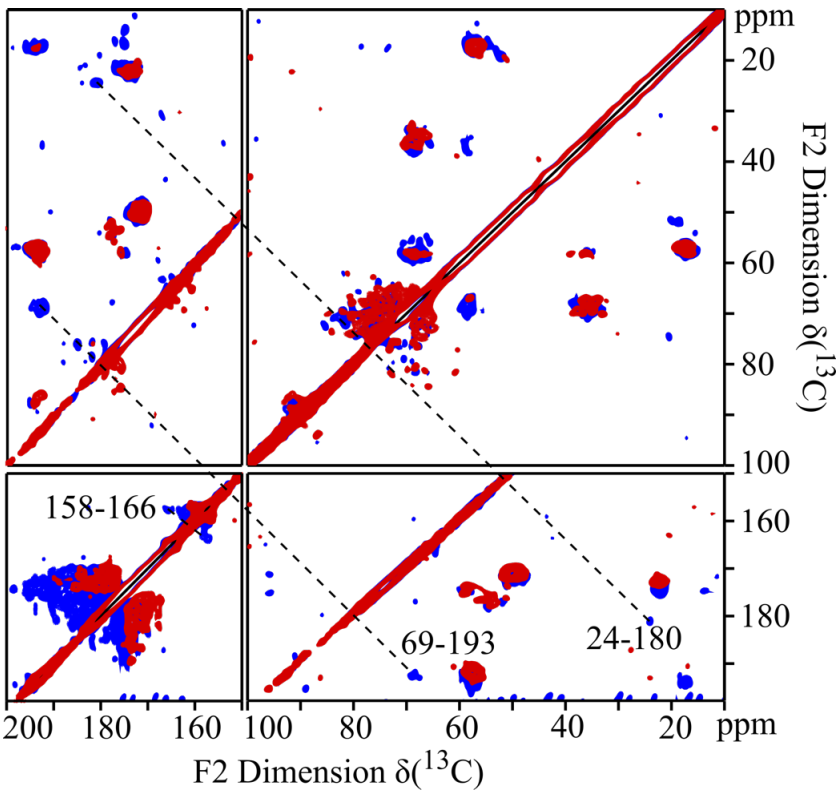


Figure 7-9 An overlay of 2D ^{13}C-^{13}C PDSD SSNMR spectra for U-$^{13}C_5$-R5P glycated collagen recorded with different mixing times. Red: 20 ms (ns = 816, 300 F1 slices). Blue: 100 ms (ns = 800, 250 F1 slices). Black dash lines indicate the correlation signals only observed with the longer mixing time.

To summarise the information we gained from this U-$^{13}C_5$-R5P glycated collagen sample, Table 7-2 includes all ^{13}C-^{13}C correlation signals observed in 2D spectra, both PC7 and PDSD, ignoring the overlapping correlations in the sugar region. Observed correlation signals are also compared with previously published data on U-$^{13}C_5$-Rib glycated collagen, and it is apparent that U-$^{13}C_5$-R5P glycated collagen contains a wider variety of glycation product signals than U-$^{13}C_5$-Rib glycated collagen. Possible assignments for the correlation signals are also suggested in Table 7-2 based on published structures and chemical shifts (some are included in Appendix).

Table 7-2 All ^{13}C-^{13}C correlation signals observed in both 2D PC7 and PDSD spectra in Figure 7-8 and Figure 7-9, and their suggested assignments based on literature (see Appendix). Chemical shift pairs underlined are also observed from U-$^{13}C_5$-Rib glycated collagen.

Cross Peak Correlations		**PC7**	**PDSD (20 ms)**	**PDSD (100 ms)**	**Suggested Assignments or Fragments**
Signal 1 /ppm	**Signal 2 /ppm**				
17	57	Yes	Yes	Yes	CEL and/or MG-H and/or MODIC
17	193	-	Yes	Yes	
19	52	-	Yes	Yes	CEL and/or MG-H and/or MODIC
22	91	Yes	-	-	**C**H_3-**C**H(OH)NH-R
22	173	Yes	Yes	Yes	N-Acetyl-
24	180	-	-	Yes	N-Acetyl-
37	59	Yes	Yes	Yes	Pentosinane and/or DOPDIC
37	69	-	Yes	Yes	DOPDIC
50	170	Yes	Yes	Yes	CML
52	177	-	Yes	Yes	
54	176	Yes	Yes	Yes	CEL and/or MG-H and/or MODIC
57	193	Yes	Yes	Yes	Norpronyl Lys
59	69	-	Yes	Yes	
59	175	-	Yes	Yes	
69	193	-	-	Yes	
88	169	Yes	-	-	
158	166	-	-	Yes	
165	165	Yes	-	-	-NH-**C**(=O)-**C**(=O)-NH-

Correlated ^{13}C pairs observed in the 2D PC7 spectrum can be interpreted as directly bonded ^{13}C pairs, and hence are expected to be observed in PDSD spectra. However, three correlated ^{13}C pairs are only observed in the PC7 spectrum, including 22-91 ppm, 88-169 ppm and 165-165 ppm. The correlation signals at 22-91 ppm observed in the 2D PC7 spectrum (Figure 7-8) is attributed to a methyl carbon ($-CH_3$), whose C-H single bonds are expected to rotate freely at room temperature. This movement can compromise the efficiency of ^{1}H assisted spin diffusions, making correlations involving methyl groups difficult to observe in PDSD experiments. In terms of the signal at about 165 ppm on F2 dimension and 330 ppm on F1 dimension in the PC7 spectrum, it is likely to be from a self-correlation between two environmentally similar and directly-bonded carbon-13 and is assigned to be oxalamide, -NH-**C**(=O)-**C**(=O)-NH-. This kind of self-correlations or correlations between structurally-similar carbon-13 is not likely to be resolved in PDSD spectra as it would fall on the diagonal like all other self-correlated carbon-13 pairs. However, looking carefully at the PDSD spectrum for this U-$^{13}C_5$-R5P glycated collagen sample in Figure 7-9, it seems that there is some off-diagonal intensity around 165-165 ppm, indicating the presence of a correlation signal. Taking into consideration the broadness of this signal, with half-height width over 3 ppm (measured by extracting slices from the 2D PC7 spectrum), the signal potentially covers slight chemical shift differences between the two correlated carbon-13 due to the special local environment they are in and their conformations. For instance, it could be a carbon-13 at 164.6 ppm correlating with another carbon-13 at 165.4 ppm. This might explain the off-diagonal intensity seen in the PDSD spectrum around 165-165 ppm.

Apart from the main signal centred at 17 ppm and 57 ppm (in both 2D PC7 and PDSD spectra), it seems that there are some weaker signals with similar chemical shifts, such as 19-52 ppm and 17-54 ppm. The ^{13}C signal at 54 ppm seems to correlate with another ^{13}C at 176 ppm (indicated by the vertical orange dotted line in Figure 7-8). They could be from CEL or MG-H or MODIC in different conformations or environments, resulting in slightly different chemical shifts for each conformation/ environment. Alternatively, these three pairs of correlated signals could be from CEL, MG-H and MODIC respectively. Considering the fact that Hyl reacts with sugar faster than Lys [78] but is present less in numbers in collagen, there could also be a contribution from CE-Hyl. All these possibilities might explain the broadness of the main 17-57 ppm correlating signal in addition to the presence of weaker signals around.

A third 2D experiment was done on this U-$^{13}C_5$-R5P glycated collagen sample, the spectrum from which, shown in Figure 7-10, supports the assignment of the oxalamide-like structure. The experiment, called the HETCOR experiment, detects the through space heteronuclear correlation between ^{1}H and ^{13}C, which is much more abundant in biological samples compared to ^{13}C and ^{13}C interactions (natural abundance of ^{1}H is around 99.99%). It also means, signals observed in this experiment are a combination of ^{13}C-enriched glycation products derived from U-$^{13}C_5$-R5P and non-isotope labelled collagen. The signal at around 165 ppm on the F2 dimension and 8.4 ppm on the F1 dimension (highlighted in a red rectangle in Figure 7-10) means this carbon-13 giving rise to a signal at 165 ppm is correlating with proton spins with chemical shifts around 8.4 ppm, which falls into the nitrogen bonded proton (**^{1}H**-N-) region. Therefore, this ^{1}H-^{13}C HETCOR spectrum in Figure 7-10 confirms that the ^{13}C at 165 ppm is close to a protonated nitrogen like what is in the oxalamide-like structure, -N**^{1}H**-**^{13}C**(=O)-**^{13}C**(=O)-N**^{1}H**-.

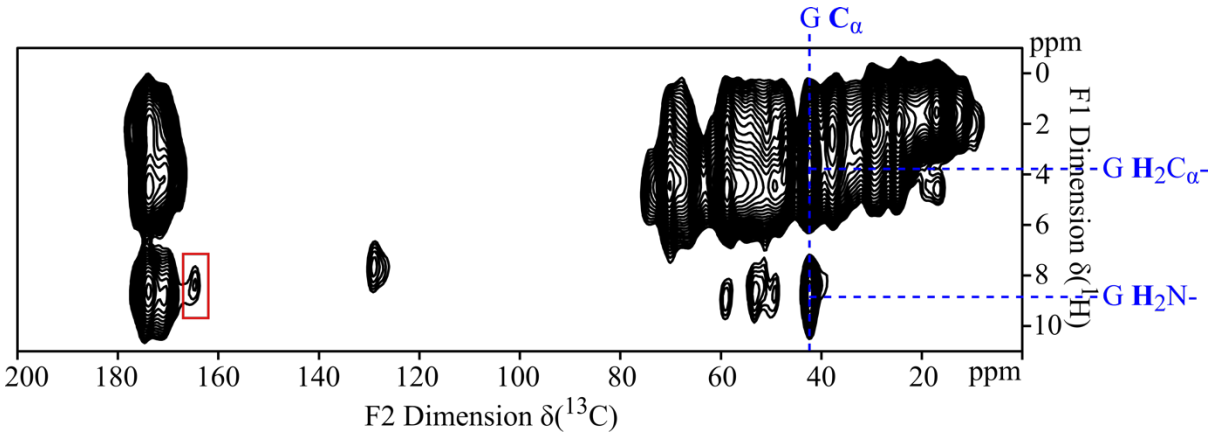


Figure 7-10 2D ^{1}H-^{13}C HETCOR SSNMR spectrum of U-$^{13}C_5$-R5P glycated collagen (ns=256 and 50 F1 slices). Correlation between ^{1}H (8.4 ppm) - ^{13}C (165 ppm) is highlighted in a red rectangle, suggesting the presence of oxalamide-like structure, -N**^{1}H**-**^{13}C**(=O)-**^{13}C**(=O)-N**^{1}H**-. ^{1}H-^{13}C correlations within G are connected by blue dash lines.

Combining information extracted from all spectra enables a better chance of assigning glycation products. A ^{13}C signal at 193 ppm, assigned as norpronyl lysine before [163], from the 2D PC spectrum, is directly bonded to a carbon-13 giving the signal at 57 ppm, and this carbon-13 bonds to a third one at 17 ppm (^{13}C connectivities indicated by solid orange lines in Figure 7-8). From the 2D PDSD spectrum, the correlation between 17 ppm and 193 ppm is observed, together indicating these three carbons are within one molecule in order of ^{13}C (17 ppm) - ^{13}C (57 ppm) - ^{13}C (193 ppm), which seems to be contradictory to the original

assignment as norpronyl lysine. Also, it is possible that this three-carbon fragment has contributions from more than one glycation product. For example, the cross peak at 17-57 ppm might be from CEL, MG-H or MODIC, and the cross peak at 57-193 ppm might be from norpronyl-lysine, and a third glycation product contains the ^{13}C (17 ppm) - ^{13}C (57 ppm) - ^{13}C (193 ppm) fragment. From 2D PC7 spectrum in Figure 7-8, it seems that there is also a three-carbon fragment, ^{13}C (173 ppm) - ^{13}C (22 ppm) - ^{13}C (91 ppm) (indicated by solid orange lines in Figure 7-8), which could be from structures like -HNO**^{13}C**-/**^{13}C**OOH-**^{13}C**(?)-**^{13}C**H(OH)NH. However, the interaction between ^{13}C (22 ppm) - ^{13}C (91 ppm) is not observed in PDSD, indicating this ^{13}C (22 ppm) is more likely to be a terminal methyl group rather than a bridging methylene group. Plus, a methylene carbon bridging between a carboxylic group or an amide group and a hydroxylated carbon is expected to have a higher chemical shift than 22 ppm. Therefore, these three signals are likely to be from two two-carbon fragments, ^{13}C (173 ppm) - ^{13}C (22 ppm) and ^{13}C (22 ppm) - ^{13}C (91 ppm), rather than the initially expected three-carbon fragment.

Assignments of all observed signals are attempted by comparing literature and experimental chemical shifts, but possible structures are not limited to the suggestions made in the above paragraphs. Because of the broadness of SSNMR signals and the similarities of glycation product structures, signals can easily overlap with one another, such as the CEL, MG-H or MODIC situation explained in the earlier paragraph. Unless a full ^{13}C connectivity within a molecule is identified, it is hard to confirm a specific structure; therefore, more possible structures are suggested based on ^{13}C chemical shifts revealed in 1D SSNMR experiments (Figure 7-11)

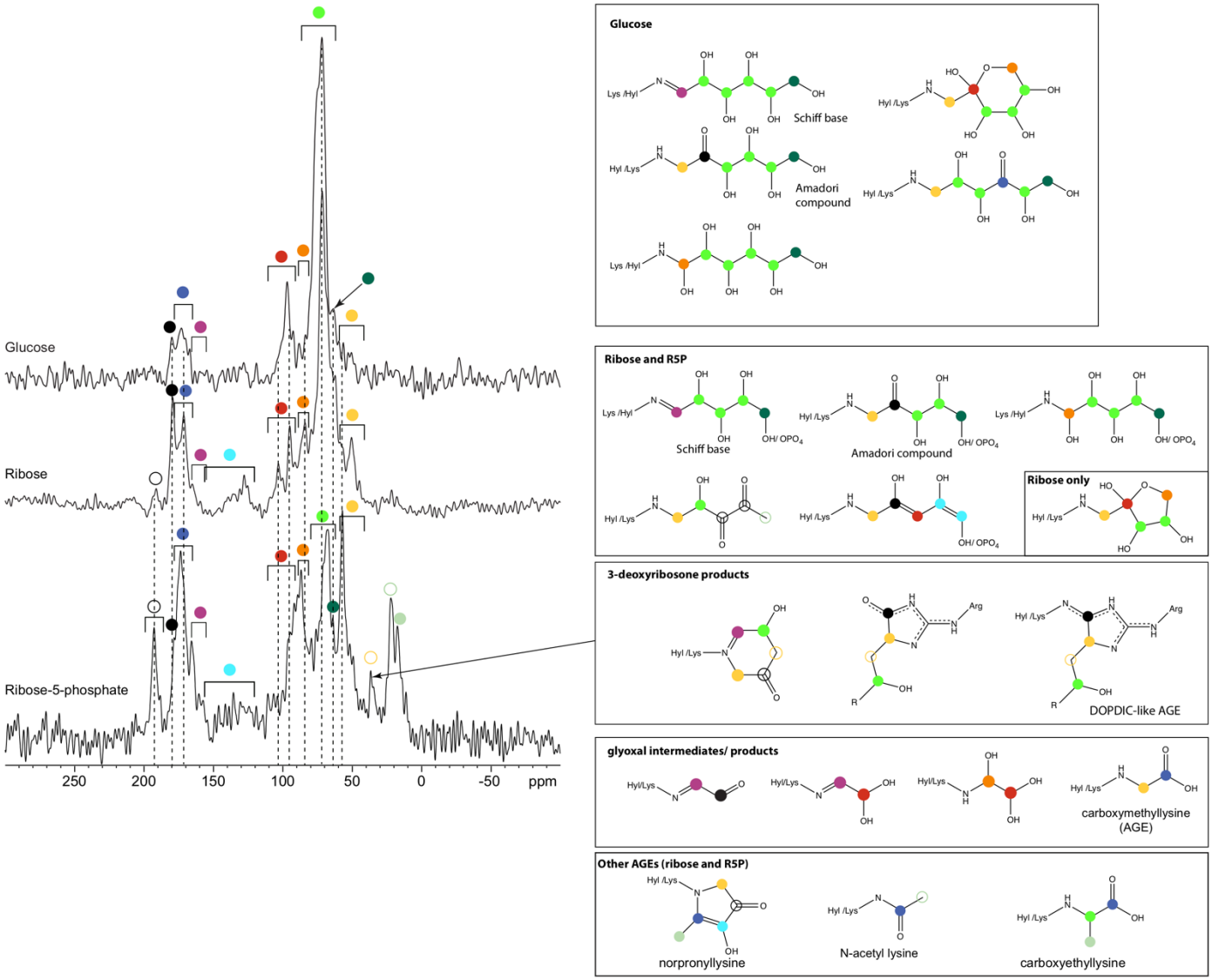


Figure 7-11 Suggested glycation products: structures and chemical shift ranges. Left: ^{13}C PC71D SSNMR spectra of U-$^{13}C_6$-Glc, U-$^{13}C_5$-Rib and U-$^{13}C_5$-R5P glycated pure collagen type I. Possible assignments for each signal or range of signals are indicated with coloured circles. Dotted lines compare chemical shifts of ^{13}C signals across three spectra. Right: possible glycation products with expected ^{13}C chemical shifts indicated with coloured circles, corresponding to the colour schemes used in the spectra on the left. Figure credit to Prof. Melinda Duer.

7.2 Glycation Products Derived from Collagen extracted from In vitro Extracellular Matrix with U-$^{13}C_6$-Glucose, U-$^{13}C_5$-Ribose-5-Phosphate and U-$^{13}C_5$-Ribose

In vitro ECM harvested from mammalian cell culture was cleaned by chymotrypsin to remove as many non-collagenous (NC) proteins as possible, leaving only collagen fibrils (more details in Chapter 6). Though the ECM primarily consists of collagen type I, other collagen types are present in minor quantities in the ECM and are resistant to chymotrypsin, and therefore are likely to remain after chymotrypsin digestion. This means compared to the pure collagen type I described in the previous section (section 7.1), samples described in this section contain a mixture of collagen types, such as type I, III, IV and V. Different collagen types are not expected to give very different glycation products, especially those monovalent amino acid sidechain modification products. The extent of glycation and glycation induced crosslinks for different collagen types might be different, depending on their unique sequences and tertiary structures. For instance, the number of sugar-accessible reactive residues and distance between two reactive residues along the collagen sequence will affect the potential to form crosslinks.

Differences between pure collagen type I and cell-culture produced collagen also arise from possibly different developmental stages of collagen, including differences in fibril sizes and maturity of enzymatic crosslinks. Observed under TEM, the diameter of pure mature collagen type I (from bovine Achilles tendon) tends to be around 100 nm, while the diameter of collagen extracted from in vitro ECM varies between 40-60 nm (images are not shown here). In other words, there is a surface area/volume (S/V) ratio difference between the two materials. Higher S/V ratio potentially exposes more reactive residues on collagen surface to ambient glycating agents, which might lead to different extents of glycation and possibly different glycation products. In terms of maturity, cell culture produced collagen is probably less mature than the pure collagen type I extracted from bovine Achilles tendon, and therefore less Hyl/Lys is enzymatically crosslinked, and more Hyl/Lys is available for glycation. In addition, early stage enzymatic crosslinks or allysine might also involve in the glycation reaction.

7.2.1 U-$^{13}C_6$-Glc Glycated Col(FSOB)

The ECM harvested from foetal sheep osteoblast (FSOB) culture was cleaned by chymotrypsin treatment and then incubated with 100 mM U-$^{13}C_6$-Glc for six months before SSNMR analysis. This prolonged incubation time was determined because the previous pure collagen type I sample made from four months incubation does not generate many AGEs signals under SSNMR.

Figure 7-12 (A) compares 1D ^{13}C CPMAS SSNMR spectra of U-$^{13}C_6$-Glc glycated pure collagen type I (the red spectrum in Figure 7-12 (A)) and collagen extracted from FSOB ECM (Col(FSOB)) (the green spectrum in Figure 7-12 (A)), in which much more intense peaks, relative to the collagen C' signal, are observed in the "sugar region" in the Col(FSOB)-Glc sample than the U-$^{13}C_6$-Glc glycated pure collagen type I sample (described in section 7.1.1), suggesting the higher extent of glycation in Col(FSOB). This could be due to the longer incubation period, less mature fibrils in Col(FSOB), greater fibril surface area/volume ratio, or a combination of these. However, not dissimilarly to the glycated collagen type I sample (the red spectrum in Figure 7-12 (A)), most glycation species generated in this sample are sugar-like intermediates.

A weak ^{13}C signal around 193 ppm is observed in the Col(FSOB)-Glc sample (the green spectrum in Figure 7-12 (A)), which is a characteristic signal of AGEs and also seen in all U-$^{13}C_6$-Glc, U-$^{13}C_5$-Rib and U-$^{13}C_5$-R5P glycated collagen. However, this signal is too weak to be observed in the 2D ^{13}C-^{13}C PDSD spectrum, shown in Figure 7-12 (B). Only ^{13}C-^{13}C correlations between hydroxylated sugar-like carbons (60-100 ppm) are detected in the 2D PDSD spectrum (Table 7-3).

The observations for this Col(FSOB)-Glc sample leads to the same conclusion as from section 7.1.1: though glucose is the most biologically relevant glycating agent, its reaction rate is too slow to be investigated as the primary glycating agent in this project. More reactive glycating agents need to be used to generate enough glycation products for SSNMR characterisation.

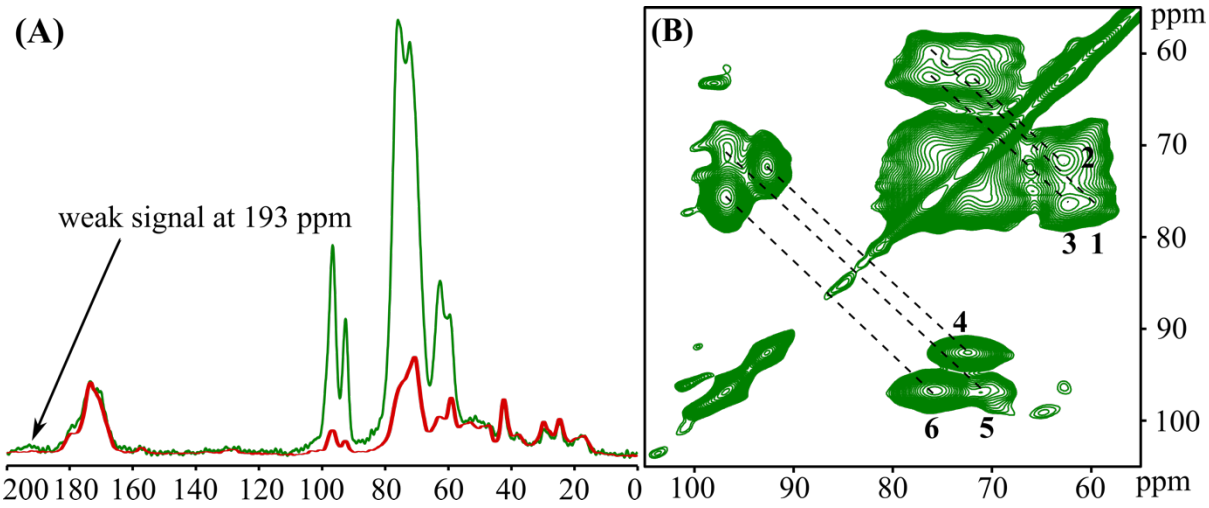


Figure 7-12 (A) An overlay of 1D ^{13}C CPMAS SSNMR spectra of U-$^{13}C_6$-Glc glycated pure collagen type I (red, ns = 5472) and collagen extracted from in vitro FSOB ECM (Col(FSOB)) (green, ns = 2048). ^{13}C signal intensity is normalised on the collagen ^{13}C' signal. The green spectrum shows a weak peak at 193 ppm (pointed by an arrow) which is absent in the red spectrum. (B) 2D ^{13}C-^{13}C PDSD SSNMR spectrum of U-$^{13}C_6$-Glc glycated Col(FSOB) (Col(FSOB)-Glc) (ns = 352 and 309 F1 slices). ^{13}C-^{13}C correlation signals are only observed in the "sugar region", and their chemical shifts are listed in the Table 7-3 below.

Table 7-3 Correlated ^{13}C-^{13}C pairs detected by 2D PDSD experiments in the Col(FSOB)-Glc sample. All signals are in the "sugar region" and are likely to be from hydroxylated carbons in glucose or glucose adducts in different conformations. ^{13}C signals at δ(C1), δ(C4) and δ(C6) of glucose are more distinctive than δ(C2), δ(C3) and δ(C5) of glucose, so, the latter three signals are not distinguished and assigned as Cx because of the uncertainty in signals assignment.

Signal Number in Figure 7-12 (B)	**Cross Peak Correlations**		**Assignments**
	Signal 1/ ppm	**Signal 2/ ppm**	
1	60	76	α-pyranose C6-C4
2	63	72	β-pyranose C6-Cx
3	63	76	β-pyranose C6-C4
4	72	93	α-pyranose Cx-C1
5	71	97	β-pyranose Cx-C1
6	76	97	β-pyranose C4-C1

7.2.2 U-$^{13}C_5$-R5P Glycated Col(FSOB) and Col(VSMC)

In vitro ECMs produced by two different cells, foetal sheep osteoblasts (FSOBs) and adult bovine vascular smooth muscle cells (VSMCs), were cleaned by chymotrypsin and then incubated with 50 mM U-$^{13}C_5$-R5P for two months separately (This concentration and incubation time apply to all samples described afterwards, unless otherwise specified). Both materials are mainly composed of collagen type I, and minor components include collagen type III, IV and more, as described above, though exact protein compositions in each ECM is unknown. Results to identify glycation products in both materials are compared in this section and also compared with U-$^{13}C_5$-R5P glycated pure collagen type I described in section 7.1.2.

1D ^{13}C CPMAS SSNMR spectra of both Col(FSOB) and Col(VSMC) are compared in (Figure 6-3 in section 6.1.3), which shows great similarities in signal frequencies and relative intensities, suggesting highly similar molecular structures and environments. So, the collagen fibrils from both sources are expected to react with R5P in a very similar manner and produce very similar glycation products. Though differences in composition between Col(FSOB) and Col(VSMC) are expected, including in their native enzymatic crosslinking, the minor collagen types that differentiate the two types of tissues they represent, i.e. bone and blood vessels, may not be abundant enough to make an observable difference in SSNMR spectra.

To characterise R5P derived glycation products and decide whether R5P reacts differently with Col(VSMC) and Col(FSOB), i.e. collagen models for two different tissues, SSNMR experiments were performed on U-$^{13}C_5$-R5P glycated Col(VSMC) and Col(FSOB) to further study glycation chemistry and products. Figure 7-13 and Figure 7-14 compare U-$^{13}C_5$-R5P glycated Col(VSMC) and Col(FSOB) by their 1D ^{13}C CPMAS spectra and 2D ^{13}C-^{13}C PDSD and PC7 spectra. No significant difference is observed in their ^{13}C 1D CPMAS SSNMR spectra (Figure 7-13), regarding ^{13}C chemical shifts of signals. Only some signal intensity variations are observed between those two samples, Col(VSMC)-R5P and Col(FSOB)-R5P (some are indicated by black arrows in Figure 7-13). These slight intensity variations could relate to the amount of a particular product generated in the two samples or different CP efficiencies, which can vary with sample conditions, such as hydration level and molecule stiffness.

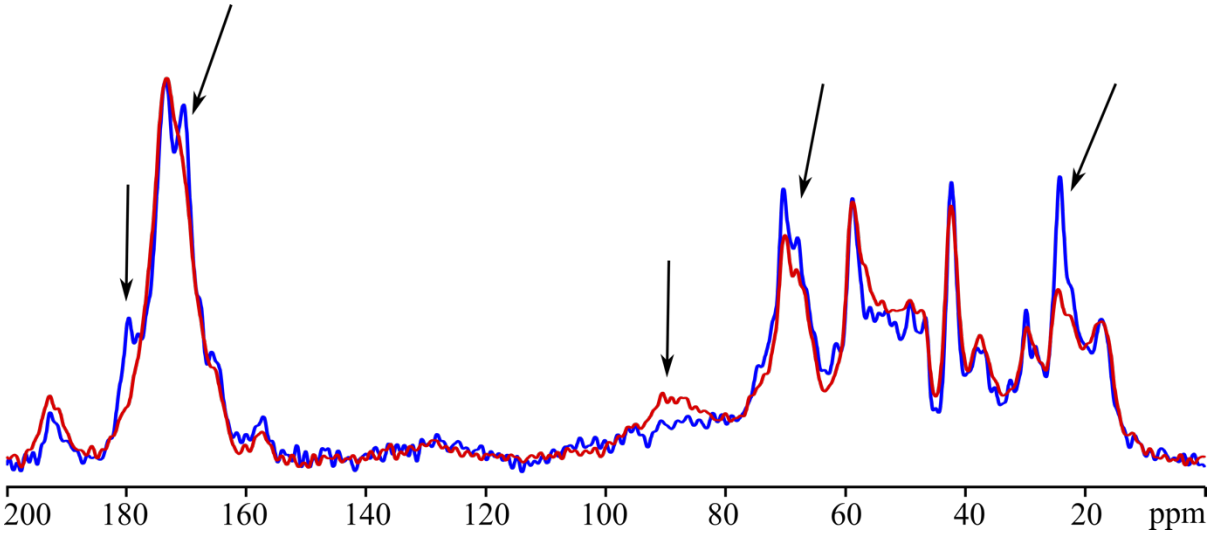


Figure 7-13 An overlay of 1D ^{13}C CPMAS SSNMR spectra of U-$^{13}C_5$-R5P glycated Col(FSOB) (blue, ns = 4096) and Col(VSMC) (red, ns = 2048), showing great similarities between them, regarding ^{13}C chemical shifts of different signals. Some signal intensities vary between the two spectra, pointed by black arrows; however, signal intensities are not the focus of this investigation.

Common correlation signal pairs observed in both samples are connected by red dash lines across both 2D PC7 and PDSD spectra (Figure 7-14), and their chemical shifts are summarised in Table 7-4. Overall, both types of 2D spectra show very similar ^{13}C-^{13}C correlation signals from ^{13}C-enriched glycation products to each other as well as to the equivalent spectrum for U-$^{13}C_5$-R5P glycated pure collagen type I. So, correlation signals that discussed in previous sections will not be repeated here.

A three-carbon fragment, ^{13}C (59 ppm) - ^{13}C (37 ppm) - ^{13}C (69 ppm), is clearly detected in the Col(VSMC)-R5P sample (Figure 7-14 A and C, indicated by green solid lines), but only the partially seen in U-$^{13}C_5$-R5P glycated pure collagen type I (will be referred to as ColI-R5P in the following text) and Col(FSOB)-R5P sample. In ColI-R5P, even though all three ^{13}C signals, at 37 ppm, 59 ppm and 69 ppm are observed in PDSD spectra with both mixing times (20 ms and 100 ms) (Figure 7-9), their connectivities are not revealed in its PC7 spectrum (Figure 7-8); while in Col(FSOB)-R5P, only the correlation between ^{13}C (37 ppm) - ^{13}C (69 ppm) is observed in the PDSD spectrum (Figure 7-14 D). The chemical shifts of this three-carbon fragment are very similar to C2, C3 and C4 in DOPDIC (structure and ^{13}C chemical shifts are shown in Appendix), suggesting the presence of DOPDIC analogue or structurally similar AGEs in R5P glycated samples.

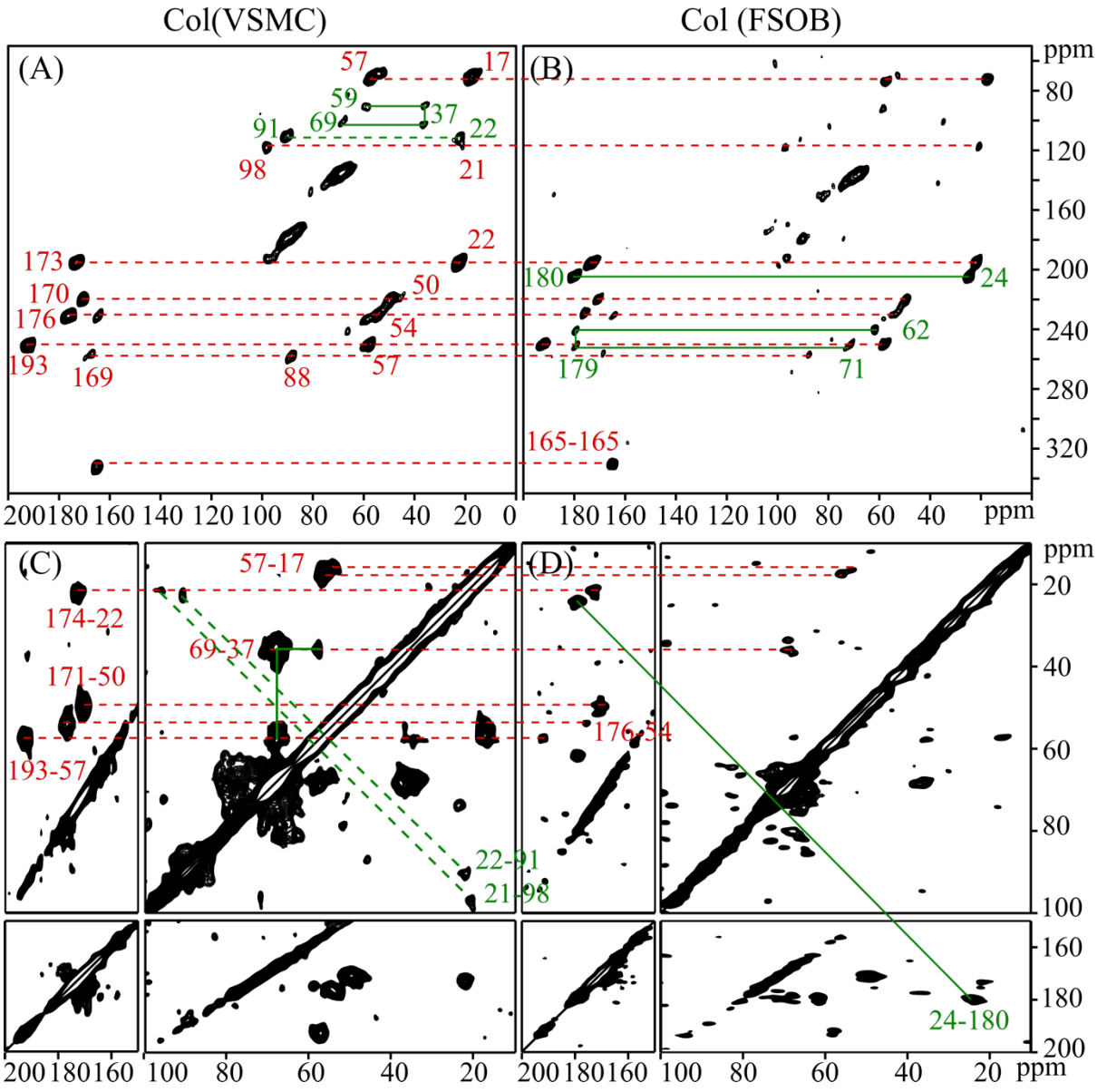
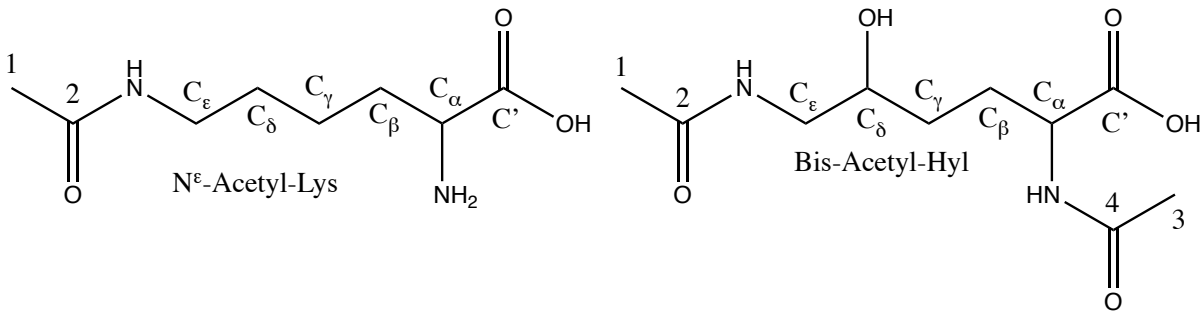


Figure 7-14 2D ^{13}C-^{13}C correlation SSNMR spectra of U-$^{13}C_5$-R5P glycated Col(VSMC) (A: PC7, ns = 2048 and 79 F1 slices. C: PDSD, 20 ms mixing time, ns = 768 and 160 F1 slices) and Col(FSOB) (B: PC7, ns = 2048 and 103 F1 slices. D: PDSD, 20 ms mixing time, ns = 624 and 319 F1 slices). Signals connected by red dash lines across spectra are common correlations observed in both samples; signals connected by green dash or solid lines mean they are observed only for one sample.

Three pairs of correlation signals are only detected in the Col(FSOB)-R5P sample: 24 ppm-180 ppm (present in ColI-R5P), 62 ppm-179 ppm and 71 ppm-179 ppm (Figure 7-14 B and D, indicated by green solid lines). The latter two could be from a hydroxylated carbon (60~80 ppm) correlating to an aldehyde carbon (~180 ppm), which could be from glycation intermediates, such as Amadori products, similar to signals **3** and **4** observed in U-$^{13}C_6$-Glc glycated pure collagen I shown in Figure 7-3. Glycation intermediates with aldehyde groups can react further and form more advanced glycation products.

The ^{13}C-^{13}C correlation signals at 24-180 ppm and 22-173 ppm are preliminary assigned to N$^{\varepsilon}$-acetyl- moieties on lysine and hydroxylysine. 1D ^{13}C solution-state NMR spectra were acquired on model compounds, N$^{\varepsilon}$-acetyl-lysine hydroxylysine and bis-acetyl-hydroxylysine (structures and experimental ^{13}C chemical shifts are shown in Figure 7-15), to measure their ^{13}C chemical shifts before confirming these assignments. It is worth noting that the 24-180 ppm correlation signal is observed in the PDSD spectrum of ColI-R5P with a mixing time of 100 ms but not the spectrum with 20 ms mixing time (Figure 7-9), which may result from the slow magnetisation transfer between methyl (24 ppm) and carbonyl (180 ppm) carbons due to the rotation of the methyl group.

Overall, the glycation products from R5P glycation of Col(FSOB) and Col(VSMC) are very similar, and they show great similarity to the products in U-$^{13}C_5$-R5P glycated pure collagen type I (described in Section 7.1.2).



Compound	C1	C3	C2	C4	C'	C_α	C_β	C_γ	C_δ	C_ε
N$^{\varepsilon}$-Acetyl-Lys	24.6	-	176.9	-	177.7	57.6	24.7	33	31	42
Bis-Acetyl-Hyl	24.5 or 24.8		177.2 or 177.3		178.7	55.8	29.8	32.8	72.4	47.8

Figure 7-15 Structures and ^{13}C chemical shifts of N$^{\varepsilon}$-acetyl-lysine and bis-acetyl-hydroxylysine measured by solution-state NMR spectra. Their ^{13}C chemical shifts (in ppm) are referenced against TSP at 0 ppm.

Table 7-4 A list of ^{13}C-^{13}C correlations observed in 2D ^{13}C-^{13}C PC7 and PDSD SSNMR spectra of U-$^{13}C_5$-R5P glycated Col(FSOB) and Col(VSMC) samples and their suggested assignments. Signals not observed for the ColI-R5P sample are in bold. [1]Observed in U-$^{13}C_6$-Glc glycated pure collagen type I in Figure 7-3.

Cross Peak Correlations		**PC7**	**PDSD (20 ms)**	**Notes**	**Suggested Assignments or Fragments**
Signal 1 /ppm	**Signal 2 /ppm**				
17	57	Yes	Yes	Both samples	CEL and/or MG-H and/or MODIC
21	**98**	Yes	Yes		
22	173	Yes	Yes		N-Acetyl-Lys
50	170	Yes	Yes		CML
54	176	Yes	Yes		CEL and/or MG-H and/or MODIC
57	193	Yes	Yes		Norpronyl Lys
88	169	Yes	-		
165	165	Yes	-		-NH-**C**(=O)-**C**(=O)-NH-
24	180	Yes	Yes	Col(FSOB)	N-Acetyl-Hyl
62	**179**	Yes	Yes		Glycation intermediates, such as Amadori products
71	**179**[1]	Yes	Yes		
22	91	Yes	Yes	Col(VSMC)	**C**H_3-**C**H(OH)NH-R
37	59	Yes	Yes		DOPDIC
37	69	Yes	Yes		DOPDIC
59	69	-	Yes		DOPDIC

7.2.3 U-$^{13}C_5$-Rib Glycated Col(FSOB) and Col(VSMC)

Ribose (structure and chemical shifts shown in Figure 7-16) is the second most abundant carbohydrate in blood, at concentrations around 100 μM [168], after glucose. Ribose, a five-carbon reducing sugar, reacts faster than glucose but slower than R5P [165]. Its glycation products with pure collagen type I [163] and amino acids [169] have been published before.

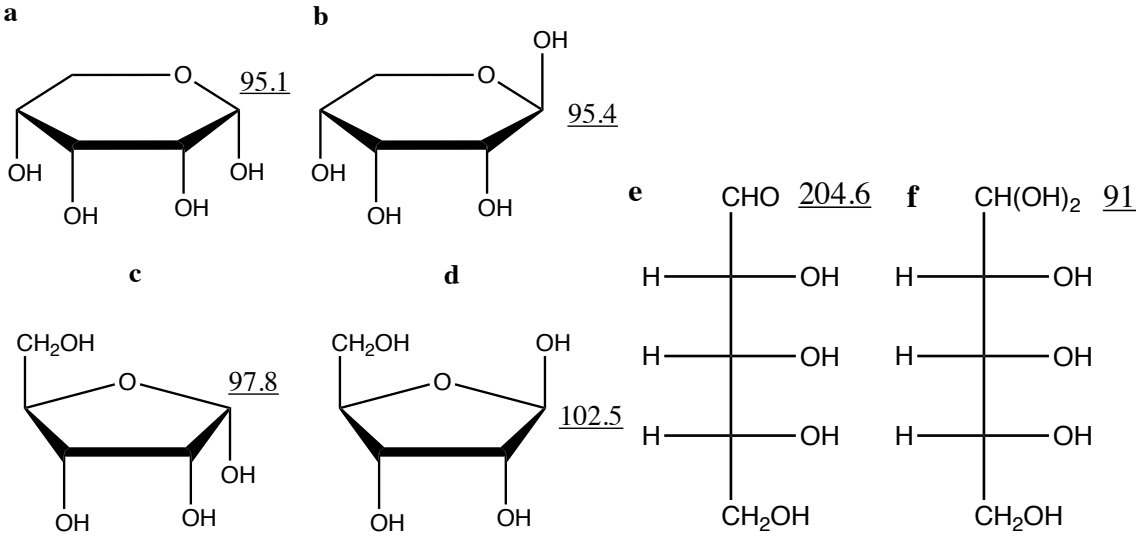


Figure 7-16 Tautomers of ribose and their C1 chemical shifts measured by ^{13}C solution-state NMR. a. α-pyranose. b. β-pyranose. c. α-furanose. d. β-furanose. e. open chain aldehyde and f. its hydrate. [170]

A ^{13}C solution-state NMR spectrum was acquired on 6-week incubated 50 mM U-$^{13}C_5$-Rib (Figure 7-17, top) to decide whether it is subject to degradation and generates similar degradation products as R5P. It is clear from the spectrum that ribose degrades during the incubation and produces fragments giving ^{13}C signals between 170-190 ppm, similar to U-$^{13}C_5$-R5P (Figure 7-17, middle). However, fragmentation products from U-$^{13}C_5$-Rib, though not assigned, are not exactly the same as those of U-$^{13}C_5$-R5P. Many sugar fragmentation products are proposed in reference [69], including different carboxylic acids.

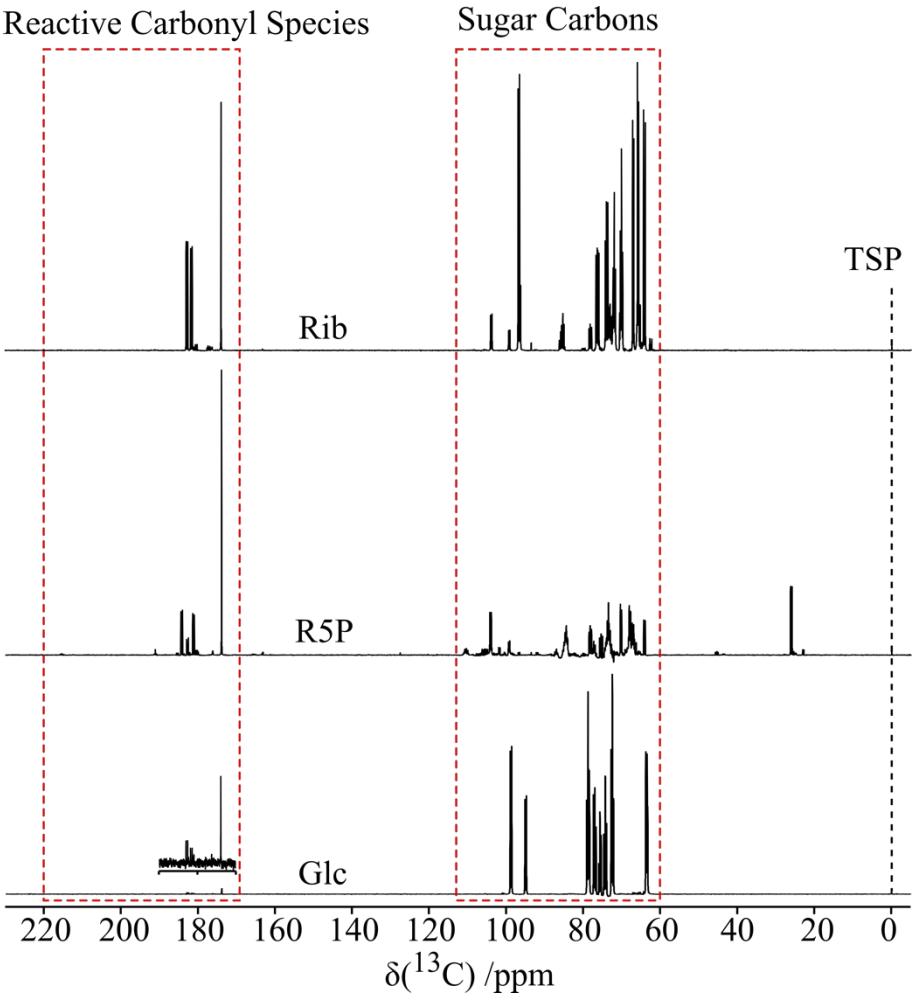


Figure 7-17 ^{13}C solution-state NMR spectra of 6-week incubated 50 mM U-$^{13}C_5$-Rib (top, ns = 1024), U-$^{13}C_5$-R5P (middle, ns = 1024) and U-$^{13}C_6$-Glc (bottom, ns = 1024) in 150 mM phosphate buffer. ^{13}C chemical shifts are referenced against TSP at 0 ppm.

Samples described in this section were prepared by incubating collagen extracted from in vitro FSOB and VSMC ECM with U-$^{13}C_5$-Rib. Different SSNMR experiments, including 1D ^{13}C CP, 2D ^{13}C-^{13}C SQ-DQ, 2D ^{13}C-^{13}C PDSD, were then performed on these two U-$^{13}C_5$-Rib glycated samples to study their glycation products and compare the glycation chemistry between Rib and R5P.

1D ^{13}C CPMAS SSNMR spectra (Figure 7-18) of both U-$^{13}C_5$-Rib glycated Col(VSMC) and Col(FSOB) samples look very similar in terms of signals arising from ^{13}C-enriched glycation products, except slight variations in relative signal intensities. The similarities in glycation products and variations in signal intensities are consistent with the observation in the corresponding spectra of the U-$^{13}C_5$-R5P glycated Col(VSMC) and Col(FSOB) samples shown

in Figure 7-13. Intense ^{13}C signals around 180 ppm and between 60-80 ppm indicate the presence of glycation intermediates. These glycation intermediate signals are resolved into a number of cross peaks in 2D ^{13}C-^{13}C correlation spectra (Figure 7-19), including cross peaks at 179-74 ppm, 179-72 ppm, 179-65 ppm and 179-62 ppm.

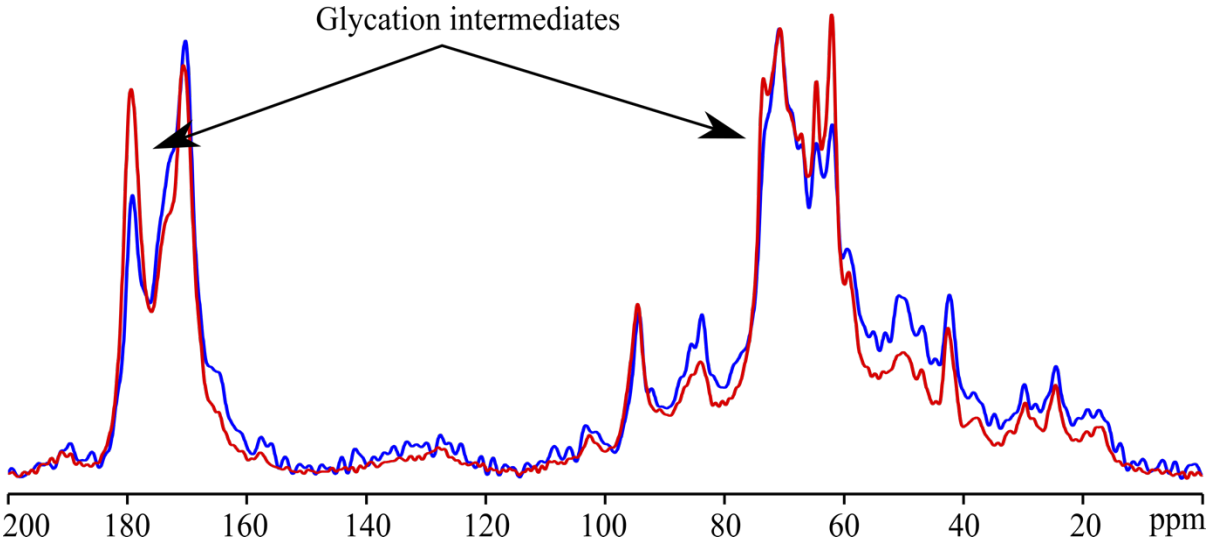


Figure 7-18 An overlay of ^{13}C CPMAS SSNMR spectra for U-$^{13}C_5$-Rib glycated Col(FSOB) (blue, ns = 4096) and Col(VSMC) (red, ns = 3072), showing intense signals assigned to glycation intermediates.

Compared with glycation products derived from U-$^{13}C_5$-R5P, most products detected for U-$^{13}C_5$-Rib glycated Col(FSOB) and Col(VSMC) appear to be very similar (see Figure 7-19 and Table 7-5). Most correlation signals observed in Figure 7-19 have been observed for previous samples before, meaning R5P and Rib generate very similar glycation products simply with different glycation rates. However, some new signals are observed in Rib glycated samples compared with R5P glycated samples, such as the correlated ^{13}C pairs at 169-169 ppm in Figure 7-19 (A) and (B), the cross peak at 51-191 ppm in Figure 7-19 (C).

Nevertheless, different products observed here, between Rib-glycated samples and R5P-glycated samples, does not necessarily mean that Rib and R5P have different glycation reaction pathways and could simply be caused by differences in their glycation rates. Rib is known to glycate more slowly than R5P, which means Rib glycated samples and R5P glycated samples are probably at different glycation stages when SSNMR spectra were acquired. Then, the difference in glycation products observed reflects the difference in glycation stages, more or less advanced glycation. In other words, in this model system, glycation products derived from Rib might be less mature/advanced than those derived from R5P. Therefore, it is hypothesised

that the ^{13}C-^{13}C correlation between 169-169 ppm might be from a structurally similar precursor (or previous phase or less mature) to a more advanced product giving rise to the 165-165 ppm correlation as well as the cross peak at 51-191 ppm versus 57-193 ppm.

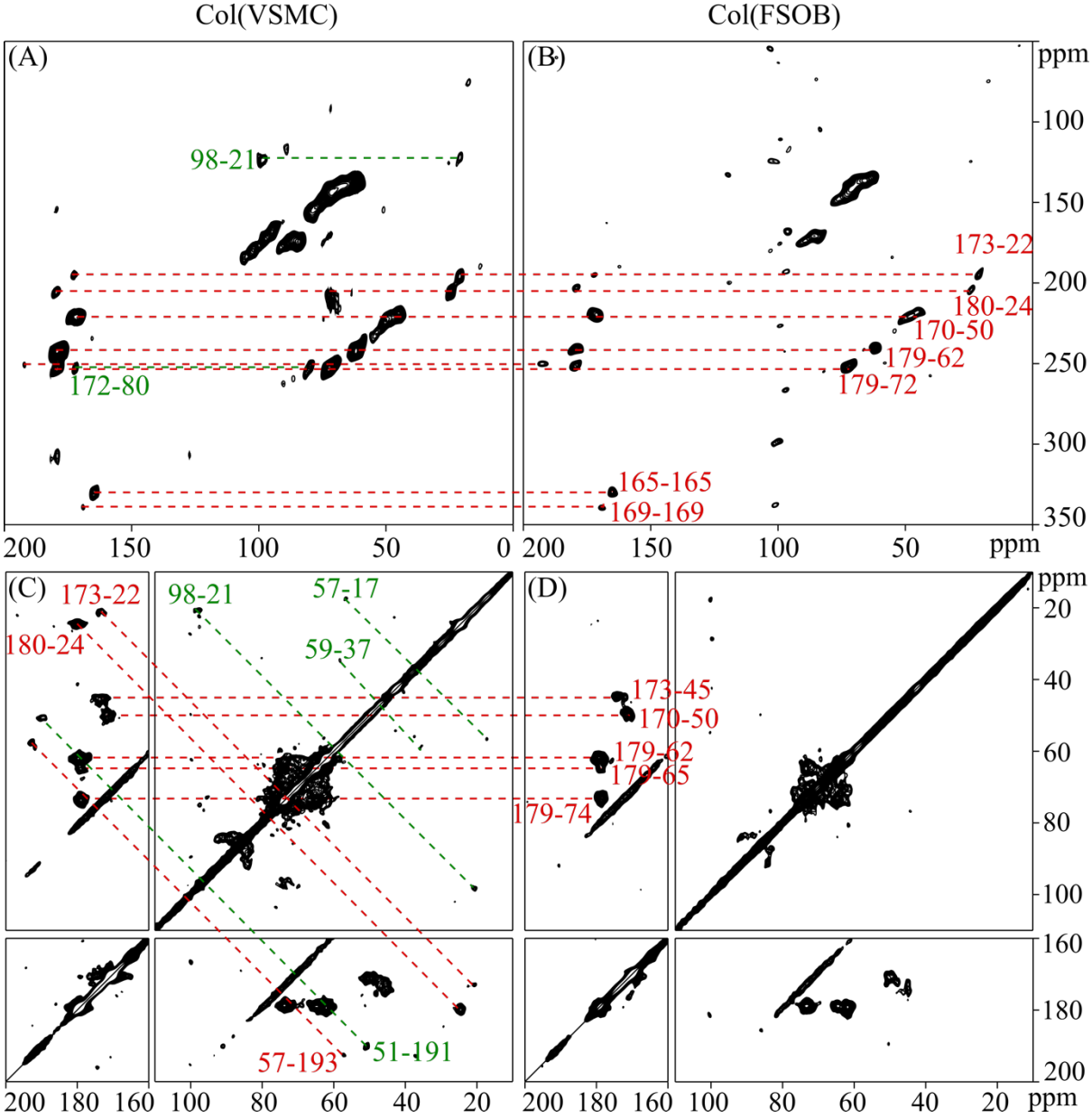


Figure 7-19 2D ^{13}C-^{13}C correlation SSNMR spectra of U-$^{13}C_5$-Rib glycated Col(VSMC) (A: PC7, ns = 1344 and 70 F1 slices. C: PDSD, 20 ms mixing time, ns = 800 and 321 F1 slices) and Col(FSOB) (B: PC7, ns = 2048 and 120 F1 slices. D: PDSD, 20 ms mixing time, ns = 1024 and 235 F1 slices). Signals connected by red dash lines are common correlations observed in both samples; signals connected by green dash lines are only observed in one spectrum.

Table 7-5 A list of correlated ^{13}C-^{13}C signal pairs observed in ^{13}C-^{13}C 2D SSNMR spectra shown in Figure 7-19 and their suggested assignments, excluding signals in the sugar region. [1]This is not well resolved and overlapped with the 50-170 ppm correlation. Signals not observed in any of the previously described samples are marked in bold.

Cross Peak Correlations		**PC7**	**PDSD (20 ms)**	**Notes**	**Suggested Assignments or Fragments**
Signal 1 /ppm	**Signal 2 /ppm**				
22	173	Yes	Yes	Both samples	N-Acetyl-Lys
24	180	Yes	Yes		N-Acetyl-Hyl
45	173	Yes[1]	Yes		
50	170	Yes	Yes		CML
57	193	Yes	Yes		Norpronyl Lys
62	179	Yes	Yes		Glycation intermediates
65	**179**	-	Yes		
72	179	Yes	-		
74	179	-	Yes		
165	165	Yes	-		-NH-**C**(=O)-**C**(=O)-NH-
169	**169**	Yes	-		
17	57	-	Yes	Col(VSMC)	CEL and/or MG-H and/or MODIC
21	98	Yes	Yes		
37	59	-	Yes		DOPDIC
51	**191**	Yes	-		
80	**172**	Yes	-		

7.3 Glycation Products Derived from In vitro Extracellular Matrix

The above sections have shown that incubating Col(FSOB) or Col(VSMC) with U-$^{13}C_5$-R5P or U-$^{13}C_5$-Rib results in very similar glycation products in pure collagen model systems, regardless of the source of collagen.

This section moves on to a more complicated model system, the integral ECM, to explore if the presence of non-collagenous (NC) proteins in the ECM would affect the glycation chemistry and generate more or different glycation products. Compared with the collagen-only systems described in the sections above, the integral mammalian ECM from in vitro cell culture has been proved to be a good in vitro model for native tissues [120] and is expected to show possible glycation products derived from NC ECM proteins.

NC proteins in the ECM, usually residing on collagen fibril surfaces to achieve biological functions, might well protect collagen in the ECM from exposure to glycating agents and thus from being glycated [171]. Also, being much smaller and more mobile than collagen, NC proteins are expected to react faster and accumulate AGEs faster than collagen. However, for those NC proteins attaching to collagen fibrils via electrostatic interactions, a consequence of glycation might be their expulsion from collagen fibrils due to charge alterations with glycation, and this could cause their precipitation out of the ECM, for instance.

VSMC and FSOB in vitro ECM were harvested by detergent lysis, and their similarities in structure and components have already been compared by their 1D ^{13}C CPMAS SSNMR spectra shown in Figure 6-3 in section 6.1.3. Four glycated in vitro ECM samples, either by U-$^{13}C_5$-R5P or U-$^{13}C_5$-Rib were prepared (see Table 7-6) and are referred to as ECM(VSMC)-Rib, ECM(VSMC)-R5P, ECM(FSOB)-Rib and ECM(FSOB)-R5P in the following text.

Table 7-6 Glycated in vitro ECM samples prepared and SSNMR experiments performed on them. [1]Also did 100 ms PDSD experiment on this sample.

Sample Type	**Sample Made**	**SSNMR Experiments Performed**					
		^{13}C CP	**^{31}P CP & DP**	**^{13}C PC7**	**^{1}H-^{13}C HECTOR**	**^{13}C-^{13}C PC7**	**^{13}C-^{13}C PDSD (20 ms)**
ECM(VSMC) – detergent lysis							
U-$^{13}C_5$-Rib Glycation	Yes	Yes	-	Yes	Yes	Yes	Yes
U-$^{13}C_5$-R5P Glycation	Yes	Yes	Yes	Yes	Yes	Yes	Yes[1]
ECM(FSOB) – detergent lysis							
U-$^{13}C_5$-Rib Glycation	Yes	Yes	-	Yes	Yes	Yes	Yes
U-$^{13}C_5$-R5P Glycation	Yes	Yes	-	Yes	Yes	Yes	Yes

1D ^{13}C CPMAS SSNMR spectra of glycated in vitro ECMs, ECM(VSMC)-Rib, ECM(VSMC)-R5P, ECM(FSOB)-Rib and ECM(FSOB)-R5P, are compared in Figure 7-20, showing similar glycation products (dash lines in Figure 7-20) and glycation intermediates signals. Glycation products accumulate in all samples and give signals between 160-200 ppm, as well as sugar-like glycation intermediates ranging between 60-90 ppm. It is also noticed that some unreacted U-$^{13}C_5$-R5P/-Rib remains in some samples. So even though four samples have been prepared and a variety of SSNMR experiments has been performed on them (Table 7-6), not all results will be presented here due to their extensive similarities to what has been observed for previous samples in the section above.

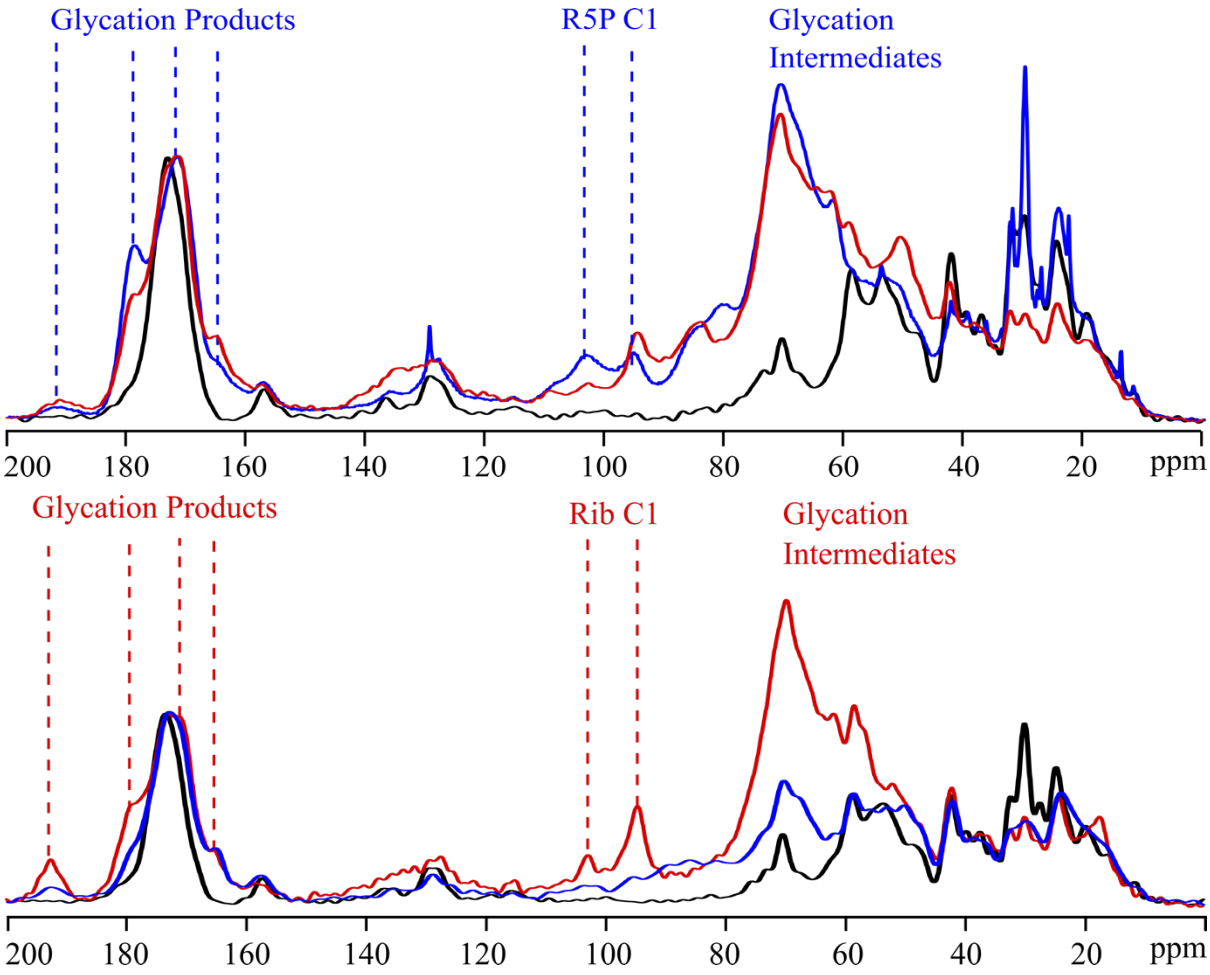


Figure 7-20 Overlays of 1D ^{13}C CPMAS SSNMR spectra of, top: ECM(VSMC) before (black, ns = 8192) and after U-$^{13}C_5$-Rib (red, ns = 10240) or U-$^{13}C_5$-R5P (blue, ns = 158928) glycation, bottom: ECM(FSOB) before (black, ns = 8192) and after U-$^{13}C_5$-Rib (red, ns = 2048) or U-$^{13}C_5$-R5P (blue, ns = 8192) glycation.

Two representative 2D ^{13}C-^{13}C correlation SSNMR spectra, PC7 and PDSD, of one sample are shown in Figure 7-21, and ^{13}C chemical shifts of correlation signals observed in the 2D spectra and possible assignments are summarised in Table 7-7. It is clear from Figure 7-21 and Table 7-7 that all signals observed in these two spectra are observed in spectra of previous samples.

Overall, glycation products detected in integral ECMs are not different from those for either collagen or ECM extracted collagen, regardless of glycating agent.

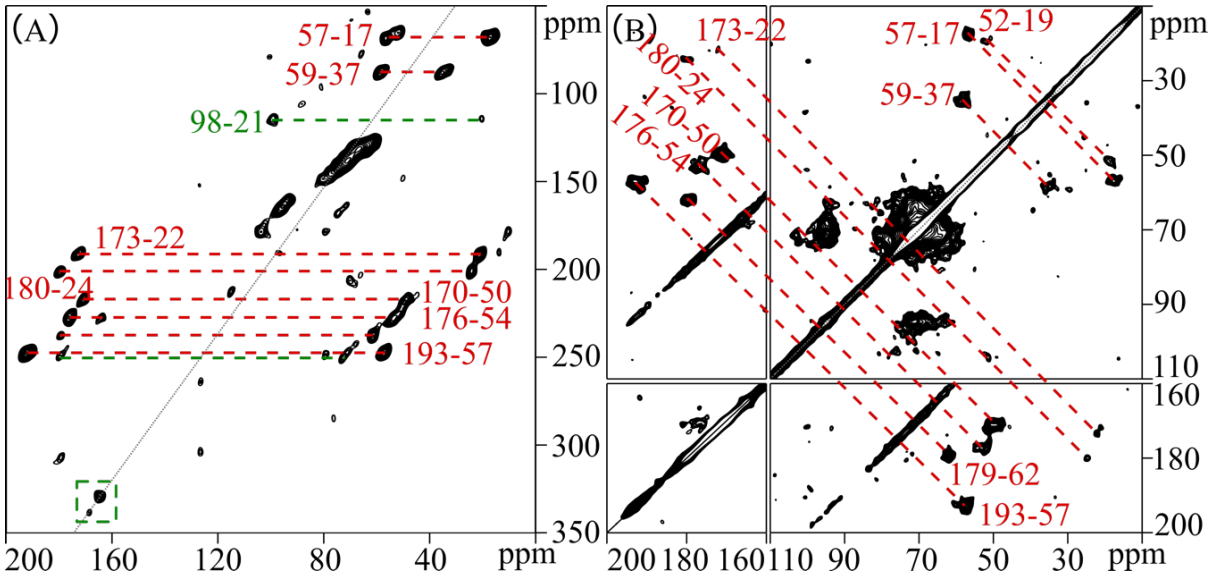


Figure 7-21 (A) 2D ^{13}C-^{13}C PC7 SSNMR spectrum (ns = 2048, 101 slices) and (B) 2D ^{13}C-^{13}C PDSD SSNMR spectrum (20 ms mixing time, ns = 896, 320 slices) of U-$^{13}C_5$-Rib glycated FSOB ECM. Correlations in red are observed in both spectra. Correlations in green are only observed in (A).

Table 7-7 A list of ^{13}C-^{13}C correlated signal pairs observed in 2D PC7 and PDSD spectra of the ECM(FSOB)-Rib sample shown in Figure 7-21 and suggested assignments for the ^{13}C-^{13}C correlation pairs.

Cross Peak Correlations		**PC7**	**PDSD (20ms)**	**Suggested Assignments or Fragments**
Signal 1/ ppm	**Signal 2/ ppm**			
17	57	Yes	Yes	CEL and/or MG-H and/or MODIC
19	52	Yes	Yes	
21	98	Yes	-	
22	173	Yes	Yes	N-Acetyl-Lys
24	180	Yes	Yes	N-Acetyl-Hyl
37	59	Yes	Yes	DOPDIC
50	170	Yes	Yes	CML and/or G-H
54	176	Yes	Yes	CEL and/or MG-H and/or MODIC
57	193	Yes	Yes	Norpronyl-Lys
62	179	Yes	Yes	Glycation intermediates
72	179	Yes	-	
165	165	Yes	-	-NH-**C**(=O)-**C**(=O)-NH-
169	169	Yes	-	

7.4 Summary

The glycation reaction is a pure chemical reaction between amino groups in proteins and aldehyde groups in reducing sugars or reducing sugar phosphates; it is associated with many diseases and widely accepted as one of the causes of various pathologies. The glycation reaction is not regulated by cell processes or enzymes, but is a pure chemistry reaction governed, as for all chemical reactions, by complex thermodynamics and kinetics, generating many possible glycation intermediates and products. In addition, glycation products evolve on both short and long timescales, from early glycation products via numerous glycation intermediates to the eventual advanced glycation endproducts (AGEs), which adds to the difficulty of following the glycation chemistry. Complicated glycation pathways, many possible glycation products and lack of cellular/enzymatic regulation in the glycation reaction pose great challenges for studying the glycation chemistry and identifying glycation products in vivo using normal biological strategies.

Therefore, in this work, the glycation reaction is studied in an in vitro model system that mimics native tissues, using solid-state NMR (SSNMR) spectroscopy in order to establish a better understanding of glycation chemistry and the major glycation products potentially generated in vivo. Three glycation substrates, pure collagen type I (from bovine Achilles tendon), collagen extracted from in vitro ECMs and intact in vitro ECMs, and three glycating agents, U-$^{13}C_6$-glucose, U-$^{13}C_5$-ribose and U-$^{13}C_5$-ribose-5-phosphate, are used in this project to study the different glycation products derived from different glycating agents. Utilisation of isotope-enriched glycating agents allows the glycation pathways to be followed from glycating agents to glycation products and the glycation products to be identified by SSNMR.

Glucose is the most bio-relevant glycating agent as it is constantly circulating in human bodies. However, it fails to accumulate a sufficient amount of AGEs for SSNMR identification, and the glucose glycation seems to be still at the reversible stage in the model system described in this work, which is less likely to cause detrimental effects compared with the formation of irreversible advanced glycation products in the case of ribose and ribose-5-phosphate glycation. Therefore, ribose and ribose-5-phosphate are used as the primary glycating agents in this work as they glycate faster than glucose and are able to accumulate a significant amount and variety of intermediate glycation products and AGEs to be characterised by SSNMR spectroscopy. The glycation chemistry and glycation products determined from ribose and ribose-5-phosphate glycation will represent at least to some extent the situation of other reducing sugars and reducing sugar phosphates respectively, since a number of other hexoses, hexose-phosphates, pentoses and pentose-phosphates are also present in human bodies as essential metabolites. However, it is possible and likely that different glycating agents generate some different glycation products and different chemical equilibria.

A number of glycation products have been identified by comparing AGE chemical shifts published in the literature and from new assignments made here by reference to model compounds, including N-acetyl-lysine and N-acetyl-hydroxylysine. Moreover, a fragment with oxalamide-like structure has been assigned, which could be derived from oxoaldehydes, such as glyoxal or methylglyoxal, originating from either fragmentation of glycation agents or glycation intermediates.

The major glycation products generated in model systems are amino acid sidechain modifications, such as N-carboxylmethyl-lysine (CML), possibly N-carboxylmethyl-arginine

(CMR), and N-acetyl-lysine/-hydroxylysine, rather than AGEs crosslinks as widely believed. Some AGEs moieties (the fragments derived from glycating agents) share very similar structures and chemical shifts, such as CEL, MG-H1 and MODIC, so it can be difficult to decide whether some ^{13}C signals are from AGE crosslinks or monovalent sidechain modifications without a complete ^{13}C connectivity map for the entire molecule, rather than just the component derived from the ^{13}C-enriched glycating agent.

Among all the glycation products identified in the model system, some are observed in almost all samples, such as CML and norpronyl-lysine, while some are less frequently observed glycation products, such as the unassigned product giving rise to ^{13}C-^{13}C correlations at 22-91 ppm (possible fragment: $^{13}CH_3$-^{13}C(OH)-N-). This variation in the glycation products observed is consistent with complex dynamics and equilibria of the glycation reaction, and some AGEs forming earlier or more readily than the others. For example, the correlated ^{13}C signals at 50-170 ppm, assigned to CML, are observed for most samples, indicating it might be one of the most facile and/ or early-forming AGEs.

Though Rib- and R5P-glycated samples share many common glycation products, a small number of glycation products are only observed in one of the two glycated samples. For example, Rib glycated samples tend to show strong ^{13}C-^{13}C correlations associated with a broad ^{13}C signal around 180 ppm and ^{13}C signals in the sugar region, as well as the ^{13}C-^{13}C correlation signal pair at 169-169 ppm in 2D SQ-DQ spectra. Possible reasons behind the differences of Rib and R5P glycation are: firstly, Rib reacts slower with amine groups than R5P, and hence, within the same incubation period, Rib probably accumulates relatively less AGEs but more glycation intermediates than R5P; secondly, R5P is negatively charged while Rib is neutral under physiological pH. This means initially R5P is more attracted to the positively-charged glycation-target residues than Rib due to charge-charge interactions; Rib can access glycated and charge-altered residues and neighbouring residues more readily due to the lack of electronic repulsion.

8 Glycation Studies in Model Systems with ^{13}C and ^{15}N Enrichment

Uniformly ^{13}C and ^{15}N enriched ECMs can be prepared by cell culture; then, ^{13}C, ^{15}N-enriched collagen can be extracted from the isotope-enriched ECMs generated. Introducing isotope enrichment of specific amino acids into the in vitro ECMs allows further studies on the glycation behaviour of specific amino acids. Moreover, ^{13}C and ^{15}N enrichment enables the performance of SSNMR experiments observing ^{15}N and ^{13}C-^{15}N correlations, to better characterise glycation products. The DNP-SSNMR technique can greatly enhance NMR sensitivity and was used on some glycated samples to detect glycation products present in minor quantities below the detection limit of conventional SSNMR. With the signal intensity enhancement of DNP, DNP-SSNMR is capable of detecting the rare Hyl and glycosylated Hyl in native ECMs [172].

Different isotope-labelling schemes allow study of different aspects of glycation. K and R are the main residues involved in the glycation reaction because of their reactive amine/guanidino groups. Therefore, in this work, U-$^{13}C_6$, $^{15}N_2$-K and U-$^{13}C_6$, $^{15}N_4$-R enriched in vitro ECMs and collagen extracted from those ECMs are glycated using U-$^{13}C_5$-R5P or U-$^{13}C_5$-Rib to gain a better understanding of the chemistry of K and R in the glycation reaction. U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P enriched ECMs are also prepared to probe the effects of glycation on collagen structure at the atomic level. When U-$^{13}C_5$, ^{15}N-P is fed into cell culture and incorporated into collagen, some U-$^{13}C_5$, ^{15}N-P are post-translationally modified into U-$^{13}C_5$, ^{15}N-O. Even though P and O do not participate directly in the glycation reaction, their chemical shifts (^{13}C and ^{15}N) are affected if the overall collagen molecular conformation is altered by the glycation reaction. Typically, an estimated isotope incorporation level of 60-70% can be achieved in in vitro ECMs (the ^{13}C incorporation level of some U-$^{13}C_6$, $^{15}N_2$-K enriched ECM samples were measured by MS; data not shown) under the cell culture conditions used in this work (details included in section 5.1). Slight variations of the isotope incorporation can occur, depending on the exact timeline of introducing isotope enrichment in relation to collagen biosynthesis by the cultured cells.

After glycation, the ^{13}C, ^{15}N-enriched collagenous materials are studied using SSNMR with a focus on studying whether ^{13}C and ^{15}N chemical shifts of enriched amino acids, K, R, G and P/O, change as an effect of glycation. For K and R, alterations in sidechain chemical environment and/ or sidechain conformation due to the attachment of glycation products at the end of their sidechains are expected to change their sidechain ^{13}C and ^{15}N chemical shifts, especially for the terminal carbons. For example, K $\delta(^{13}C_{\varepsilon})$ goes up 7 ppm from K to CML [173]; K $\delta(^{15}N_{\varepsilon})$ is also observed to be higher after glycation due to transition from a primary amine to a secondary amine. G and P/O are not directly involved in glycation reactions and so their ^{13}C and ^{15}N chemical shifts can be expected to report on any overall secondary structure changes to the collagen molecules as a result of the glycation reaction or AGE accumulation. In addition, the strength of the internal hydrogen bonding between G and X-position residues in neighbouring chains that stabilises the collagen triple helix is expected to be reflected in the G N-H ^{1}H chemical shift, which here is measured in 2D ^{1}H-^{13}C correlation SSNMR experiments.

8.1 Glycation Products Derived from U-$^{13}C_6$, $^{15}N_2$-K Enriched Model Systems with U-$^{13}C_5$-Ribose-5-Phosphate and U-$^{13}C_5$-Ribose

Lysine (K) is of special interest in studying glycation reaction, as it has a free amino group at the end of its sidechain. This free amino group of K potentially acts as a nucleophile to initiate the glycation reaction.

There are 107 K residues in the bovine collagen type I sequence (from UniProt), some of which are expected to be post-translationally modified to 5-hydroxylysine (Hyl) and then glycosylated (^{13}C and ^{15}N chemical shifts of K and post-translationally modified K are included in Table 8-1 and Table 8-2). The extent of K post-translational modifications (PTMs), hydroxylation and glycosylation, changes with animal age and is different in different tissues.

K glycation is dependent to some degree on the extent of K PTMs. As the collagen fibril matures, the extent and nature of enzymatic crosslinks also alter, which can also affect K glycation. For example, if collagen fibrils are less enzymatically crosslinked, more K will be available for glycation. We hypothesise that this might be the case for collagen produced in vitro in cell culture. However, it is worth mentioning that some latest research has shown that K glycation might interfere with the existing (reversible) enzymatic crosslinks (unpublished group data, not shown here).

Table 8-1 ^{13}C and ^{15}N chemical shifts for lysine (K), proline (P), hydroxyproline (O) and glycine (G) measured in collagenous samples in this work.

AA	δ(^{13}C)/ ppm						δ(^{15}N)/ ppm	
	C'	C_α	C_β	C_γ	C_δ	C_ε	N'	N
K	173	53	30	23	27	40	122	32
P	174	58	30	25	47	-	130	-
O	172	58	38	70	55	-	130	-
G	169	42	-	-	-	-	106	-

First, a set of U-$^{13}C_6$, $^{15}N_2$-K enriched samples were prepared, including Col(VSMC-K), Col(FSOB-K), ECM(VSMC-K) Detergent, ECM(FSOB-K) Detergent, ECM(VSMC-K) Freeze-thaw and ECM(FSOB-K) Freeze-thaw (nomenclature explained in Figure 8-1 caption), to compare the similarities and differences of their K environments and conformations in these collagenous samples. Then, those samples were glycated to study the chemistry of K glycation.

Figure 8-1 compares 1D ^{13}C CPMAS SSNMR spectra of differently treated U-$^{13}C_6$, $^{15}N_2$-K labelled in vitro ECMs, including intact in vitro ECM harvested from freeze-thaw and detergent cell lysis and collagen extracted from detergent lysed in vitro ECM using chymotrypsin to remove non-collagenous (NC) proteins. Overall, no significant difference in K ^{13}C chemical shifts is found between the different samples (indicated by the red dash lines in Figure 8-1), which aligns with the conclusion drawn from the previous chapter: molecular conformations in these collagenous samples are very similar. One general trend observed in Figure 8-1 is that when K-containing NC proteins are removed from the ECMs by chymotrypsin digestion, ^{13}C signals from natural abundant G and P/O start to be observed in the SSNMR spectra (signals indicated by black dash lines Figure 8-1). In addition, some ^{13}C signals are narrower, such as the K ^{13}C' and K $^{13}C_\alpha$, indicating that the distribution of K conformations or environments in the samples are less heterogeneous after chymotrypsin digestion. In theory, some U-$^{13}C_6$, $^{15}N_2$-K residues in collagen should be post-translationally modified to U-$^{13}C_6$, $^{15}N_2$-Hyl; however, Hyl in collagen is too rare to be observed by conventional SSNMR, and its signature signal, Hyl C_δ, expected ~ 70 ppm, potentially overlaps with O C_γ (~ 70 ppm).

Due to the similarities between Col(VSMC-K) and Col(FSOB-K) regarding K composition and conformation, results from only Col(FSOB-K) are described and discussed in the following sections, unless otherwise stated.

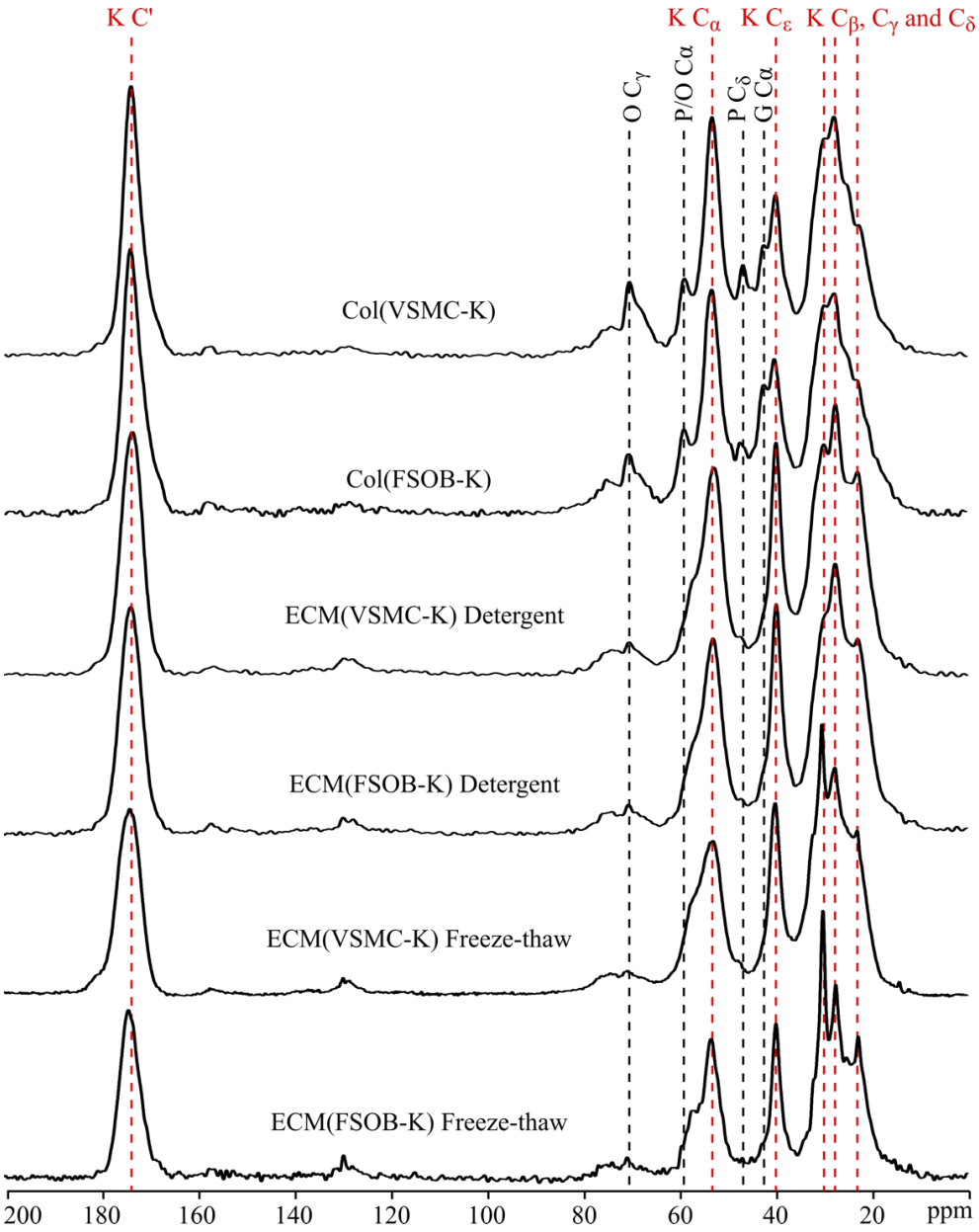


Figure 8-1 Comparison of 1D ^{13}C CPMAS SSNMR spectra of, from top to bottom, Col(VSMC-K) (U-$^{13}C_6$, $^{15}N_2$-K enriched VSMC ECM, detergent lysed, cleaned up by chymotrypsin and Tween-20, ns = 8192), Col(FSOB-K) (U-$^{13}C_6$, $^{15}N_2$-K enriched FSOB ECM, detergent lysed, cleaned up by chymotrypsin and Tween-20, ns = 4096), ECM(VSMC-K) Detergent (U-$^{13}C_6$, $^{15}N_2$-K enriched VSMC ECM, detergent lysed, ns = 4027), ECM(FSOB-K) Detergent (U-$^{13}C_6$, $^{15}N_2$-K enriched FSOB ECM, detergent lysed, ns = 4096), ECM(VSMC-K) Freeze-thaw (U-$^{13}C_6$, $^{15}N_2$-K enriched VSMC ECM, freeze-thaw lysed, ns = 36666) and ECM(FSOB-K) Freeze-thaw (U-$^{13}C_6$, $^{15}N_2$-K enriched FSOB ECM, freeze-thaw lysed, ns = 2048). Red dash lines indicate all the K ^{13}C signals, and black dash lines indicate the natural abundant ^{13}C signals from O C_γ, P/O C_α (not resolved), P C_δ and G C_α.

8.1.1 Col(FSOB-K) Glycated with U-$^{13}C_5$-R5P and U-$^{13}C_5$-Rib

Glycated U-$^{13}C_6$, $^{15}N_2$-K enriched collagenous samples are referred to as Col(FSOB-K)-R5P or Col(FSOB-K)-Rib in the following text. Figure 8-2 compares 1D ^{15}N (top) and ^{13}C (bottom) CPMAS SSNMR spectra of Col(FSOB-K)-R5P and Col(FSOB-K)-Rib with non-glycated Col(FSOB-K).

When glycation products attach to K ^{15}N, a subset of primary amine becomes secondary amine or amide, depending on the adducts, and therefore, the glycated K ^{15}N signal is expected to shift towards higher chemical shifts than its original K ^{15}N signal. This change from K to glycated K, i.e. primary amine to amide-like structure, possibly explains the broad ^{15}N signal appearing around 105 ppm after glycation (Figure 8-2 top, pointed by an arrow).

Similar ^{13}C signals from glycation products are observed in both Rib and R5P glycated samples as shown in Figure 8-2 (bottom) to those in the samples discussed in the previous chapter, as expected, such as the ^{13}C signals at 193 ppm, 180 ppm, 165 ppm and the many overlapping signals in the sugar region. It is also noticed from 1D ^{13}C CPMAS SSNMR spectra that the U-$^{13}C_5$-Rib glycated sample (the green spectrum in Figure 8-2 bottom) gives stronger (relative to the K C' signal) and a more diverse range of sugar-like glycation intermediate ^{13}C signals than the U-$^{13}C_5$-R5P glycated sample (the red spectrum in Figure 8-2 bottom). The strong ^{13}C peak around 180 ppm for the U-$^{13}C_5$-Rib glycated sample (green spectrum in Figure 8-2 bottom) also indicates the presence of reactive carbonyl species, possibly from Amadori products or other glycation intermediates. This broad ^{13}C signal at 180 ppm is resolved into several ^{13}C-^{13}C correlation signals in 2D ^{13}C-^{13}C correlation SSNMR spectra (Figure 8-4 (B) and (D) in orange), which shows correlations between ^{13}C signals at around 180 ppm and ^{13}C signals ranging from 56 ppm to 73 ppm. These correlations could come from carbonyl carbons and hydroxylated carbons in sugar adducts. This observation of relatively strong and greater diversity of ^{13}C signals from glycation intermediates in the Rib-glycated sample than the R5P-glycated sample is likely due to their differences in glycation rates, as previously discussed.

The change from K to glycated K is expected to shift a portion of the K $^{13}C_\varepsilon$ (40 ppm) signal to higher chemical shifts or broaden the K $^{13}C_\varepsilon$ signal because of the increase in heterogeneity in $^{13}C_\varepsilon$ environments; however, the shifting or broadening of K $^{13}C_\varepsilon$ (40 ppm) signal can be

difficult to extract from 1D ^{13}C spectra (Figure 8-2 bottom) because K $^{13}C_{\varepsilon}$ overlaps with natural abundance G C_{α} signals (42 ppm) and potentially with ^{13}C-enriched glycation products. For example, previously observed ^{13}C-^{13}C correlation signals (in non-isotope-enriched model system), 54-176 ppm (CEL and/or MG-H and/or MODIC) and 22-173 ppm can overlap K $^{13}C_{\alpha}$ (53 ppm) - ^{13}C' (174 ppm) and $^{13}C_{\delta}$ (23 ppm) - ^{13}C' (174 ppm), and hinder the possible chemical shift changes or signal broadening of K ^{13}C signals due to glycation.

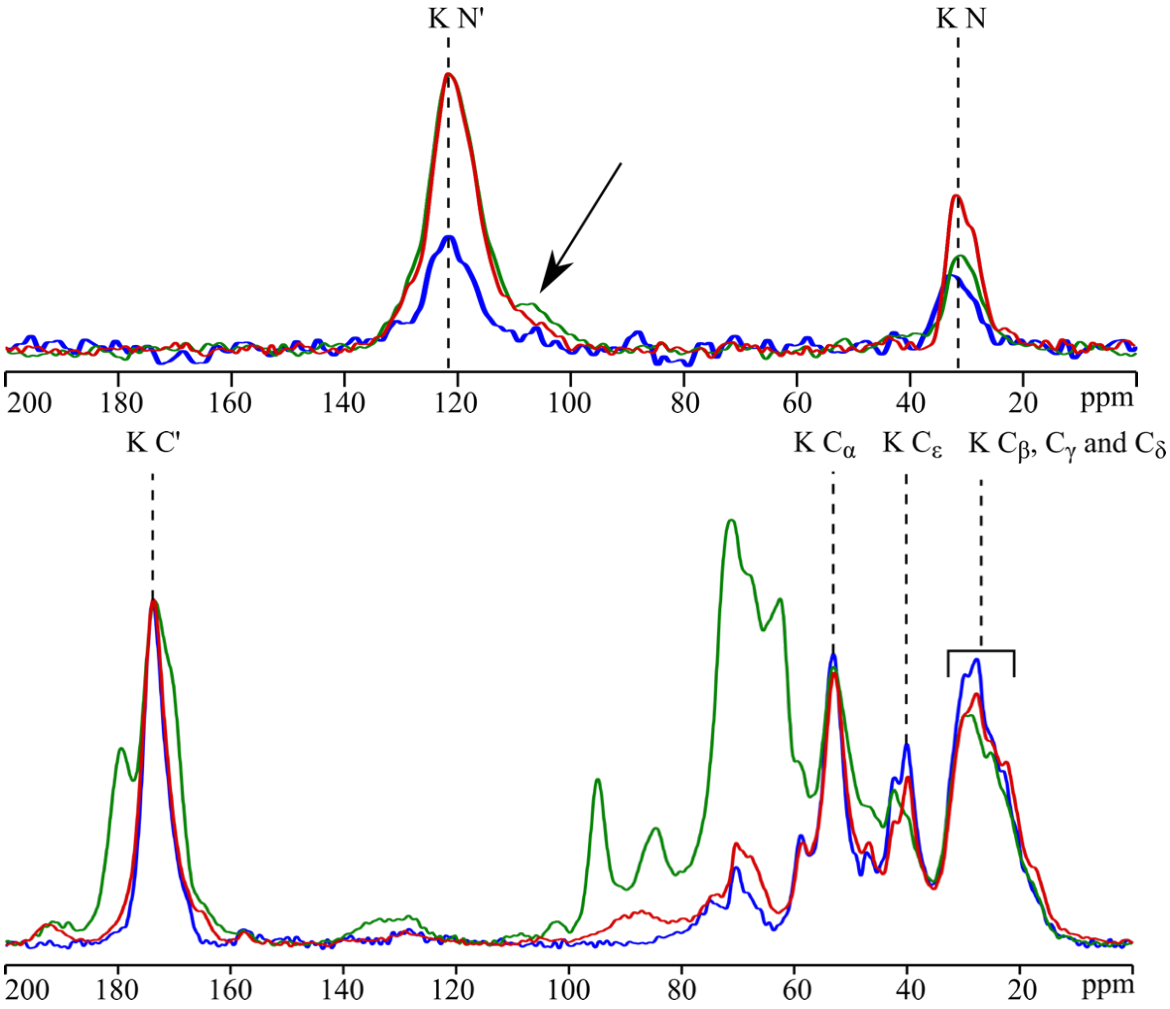


Figure 8-2 Top: 1D ^{15}N CPMAS SSNMR spectra of Col(FSOB-K) (blue, ns = 8192), U-$^{13}C_5$-Rib glycated (green, ns = 38633) and U-$^{13}C_5$-R5P glycated (red, ns = 6201) Col(FSOB-K). Bottom: 1D ^{13}C CPMAS SSNMR spectra of Col(FSOB-K) (blue, ns = 4096), U-$^{13}C_5$-Rib glycated (green, ns = 4096) and U-$^{13}C_5$-R5P glycated (red, ns = 2048) Col(FSOB-K). Signal intensities are normalised on the K C' signal. Black dash lines indicate the successful enrichment of ^{13}C and ^{15}N in K residues in these samples.

The 2D PDSD SSNMR spectrum of Col(FSOB-K)-R5P in Figure 8-4 (C) show ^{13}C-^{13}C correlations between ^{13}C (30 ppm), ^{13}C (45 ppm), ^{13}C (53 ppm), ^{13}C (68 ppm) and ^{13}C (75 ppm) (the cross peaks included in a purple dash rectangle). These ^{13}C chemical shifts are very similar to those of previously reported PTMs of K signals, Hyl, Hyl-NH-R and GlcGal-Hyl, by DNP-SSNMR (their ^{13}C chemical shifts are summarised in Table 8-2) [172]. However, Hyl composes less than 2% of K residues in collagen, and only a small portion of Hyl is further modified into Hyl-NH-R and GlcGal-Hyl. Thus, PTMs of K are unlikely to be observed by conventional SSNMR and have never previously been observed in non-glycated samples without DNP. Thus, we instead assign these ^{13}C correlation signals to glycated K and hence provide direct evidence on how ^{13}C signals shift when K becomes glycated K to form K-NH-sugar structures. A schematic of a K glycation molecular structure that can account for these ^{13}C signals is shown in Figure 8-3.

Table 8-2 ^{13}C chemical shifts (in ppm) of three post-translation modifications (PTMs) of K, Hyl, Hyl-NH-R and GlcGal-Hyl in collagen [172]. Chemical shifts in purple indicates signals with similar chemical shifts are observed in the 2D ^{13}C-^{13}C spectra of ECM(VSMC-K)-R5P in Figure 8-6.

PTMs of K	**C_α**	**C_β**	**C_γ**	**C_δ**	**C_ε**
Hyl-NH_2	54	28	30	69	45
Hyl-NH-R	54	30	28	71	56
(Glc)Gal-Hyl	54	28	30	75	44

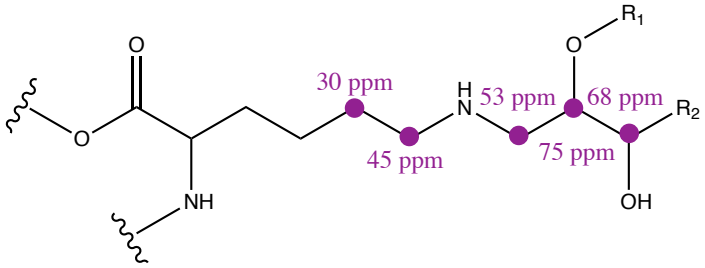


Figure 8-3 A suggested structure of a K glycation intermediate/product giving correlation signals that include the observed signals of ^{13}C (30 ppm), ^{13}C (45 ppm), ^{13}C (53 ppm), ^{13}C (68 ppm) and ^{13}C (75 ppm), i.e. the signals marked in purple or included in the purple dash rectangle in Figure 8-4. ^{13}C correlation signals can be either from U-$^{13}C_6$, $^{15}N_2$-K or U-$^{13}C_5$-Rib/-R5P in these samples where both glycation substrate (K) and glycating agents are ^{13}C-enriched. ^{13}C chemical shifts for R_1 and R_2 are not resolved in the 2D spectra.

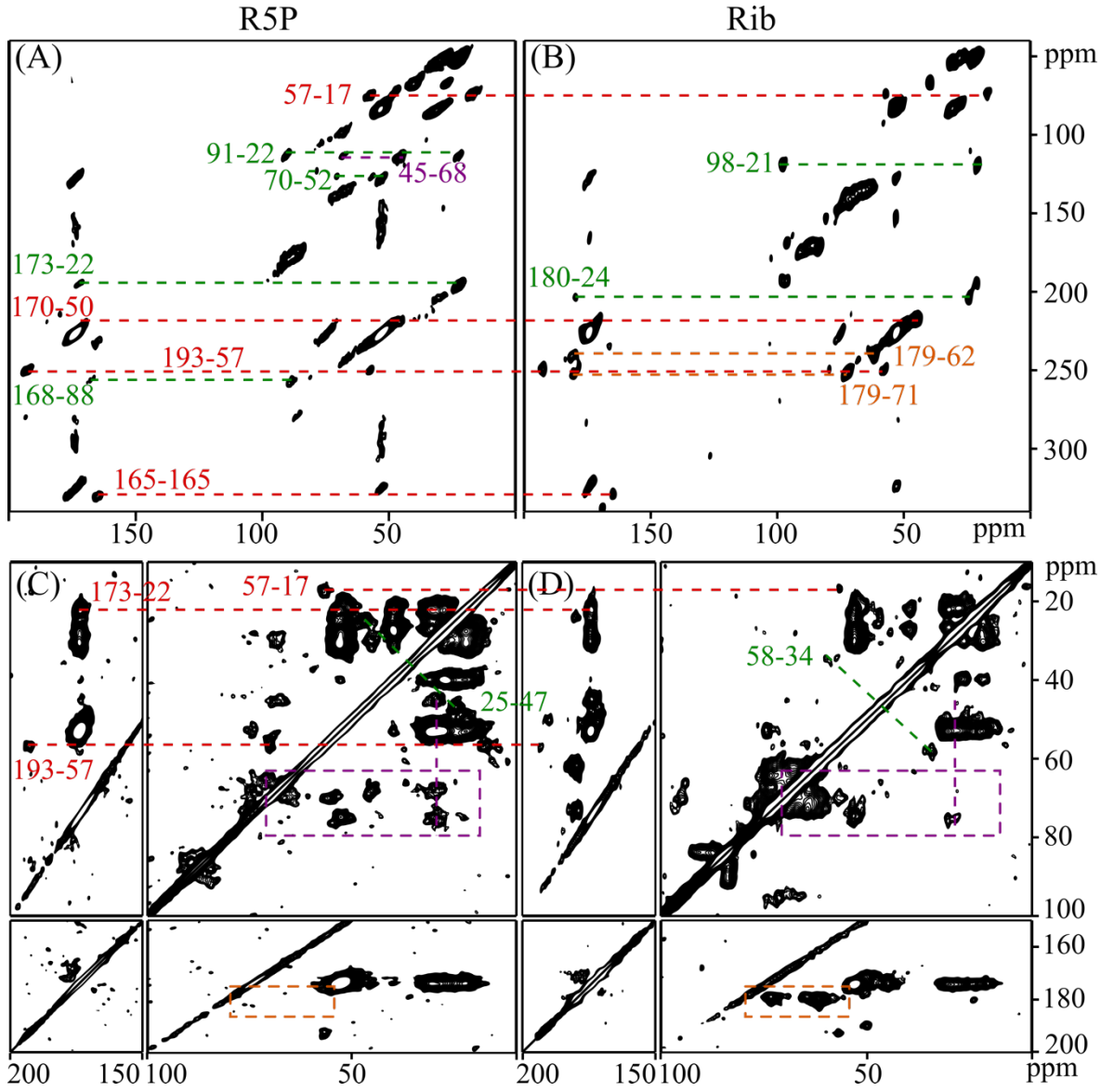


Figure 8-4 2D ^{13}C-^{13}C PC7 (top row) and PDSD (bottom row) SSNMR spectra of Col(FSOB-K)-R5P (left column; A: ns = 2048 and 120 F1 slices; C: ns = 800 and 266 F1 slices) and -Rib (right column; B: ns = 3424 and 60 F1 slices; D: ns = 912 and 180 F1 slices). Correlation pairs observed in all samples are linked by red dash lines; correlating signals only observed in one spectrum are linked by green dash lines. Purple dash rectangle shows signals arising from Hyl-like and glycosylated Hyl-like structures discussed above. Orange dash rectangle includes all signals correlated a broad 1D signal maxima at 180 ppm discussed in the text.

Table 8-3 A summary of ^{13}C-^{13}C correlation signals observed in 2D PC7 and PDSD spectra of Col(FSOB-K)-R5P and Col(FSOB-K)-Rib in Figure 8-4, excluding the overlapping signals in the sugar region and in the so-called glycation intermediates region (the orange rectangles). Chemical shifts in purple correspond to signals in the purple dash rectangles in Figure 8-4. Chemical shifts in bold are correlation signal pairs that have not been observed in any previous spectra; they might correspond to correlations between K sidechain and glycation adducts.

^{13}C-^{13}C Correlations		**R5P**		**Rib**		**Suggested Assignments or Fragments**
Signal 1 / ppm	**Signal 2 / ppm**	**PC7**	**PDSD (20 ms)**	**PC7**	**PDSD (20 ms)**	
17	57	Yes	Yes	Yes	Yes	CEL and/or MG-H and/or MODIC
21	98	-	-	Yes	-	
22	90	Yes	-	-	-	
22	173	Yes	Yes	-	Yes	N-Acetyl-Lys
24	180	-	-	Yes	Yes	N-Acetyl-Hyl
25	**47**	**-**	**Yes**	**-**	**-**	
34	58	-	-	-	Yes	
50	170	Yes	Yes	Yes	Yes	CML
51	191	-	-	-	Yes	
52	**70**	**Yes**	**-**	**-**	**-**	
57	193	Yes	Yes	Yes	Yes	Norpronyl-Lys
59	174	-	Yes	-	-	
88	**168**	**Yes**	**-**	**-**	**-**	
165	165	Yes	-	Yes	-	-NH-**C**(=O)-**C**(=O)-NH-
169	169	-	-	Yes	-	
30	45	-	Yes	-	-	See the suggested structure in Figure 8-3
30	68	-	Yes	-	-	
30	75	-	Yes	-	Yes	
45	68	Yes	Yes	-	-	
53	75	-	Yes	-	Yes	
56	70	-	Yes	-	Yes	
66	70	-	Yes	-	Yes	

8.1.2 ECM(VSMC-K)-R5P and Col(VSMC-K)-Rib: A DNP-SSNMR Study

SSNMR spectroscopy is not a highly sensitive technique and thus signals from low abundance glycation species may not be observed. Low abundance of a glycation species does not mean those species are not important; indeed, we expect highly reactive intermediates – which are clearly important in the glycation reaction scheme – to be relatively low abundance. Similarly, glycation crosslinks, which have the potential to cause significant molecular structure changes as well as changes to molecular elasticity, are expected to be very low in abundance. DNP-SSNMR generates significant sensitivity enhancement that allows observation of low abundant glycation products not observable by conventional SSNMR. More importantly, DNP-SSNMR allows acquisition of 2D ^{13}C-^{15}N correlation spectra in relative short experiment time, which were not performed on conventional SSNMR due to the excessive length of these experiments arising from the low sensitivity of ^{15}N in NMR spectroscopy. However, sacrifice is made on spectral resolution because the low operating temperature of DNP (100 K) can cause significant line broadening (explained in section 4.5). For example, the ^{13}C signals from K sidechain carbons, C_β, C_γ and C_δ, in the 1D DNP-SSNMR spectra of collagenous samples in Figure 8-5 (A), are not as well resolved as in the spectra acquired at room temperature in Figure 8-1 (the third one from top). Figure 8-5 shows that all three signals "merge" into one broad signal under DNP conditions (Figure 8-5 (A)), and it seems that the resolution on these three sidechain carbons is further degraded after glycation (Figure 8-5 (B) and (C)). Discussion in this section will focus on signals not observed by conventional SSNMR.

1D 13C DNP-SSNMR spectra of Col(VSMC-K)-Rib and ECM(VSMC-K)-R5P (Figure 8-5) show that in addition to the ^{13}C glycation signals observed before, such as the ^{13}C signals at 193 ppm, 165 ppm and the overlapping signals in the sugar region, DNP-SSNMR captures new glycation product signals in the ^{13}C chemical shifts region between 100-150 ppm than conventional SSNMR (Figure 8-5 (B) and (C)), which is where some of the aromatic crosslinks signals are expected. However, few correlation signals are observed in this region in 2D ^{13}C-^{13}C spectra (Figure 8-6 and Figure 8-7), suggesting that the glycation products giving ^{13}C signals between 100-150 ppm are either in really low quantities or not ^{13}C enriched or not bonded to a ^{13}C or have few ^{1}H around.

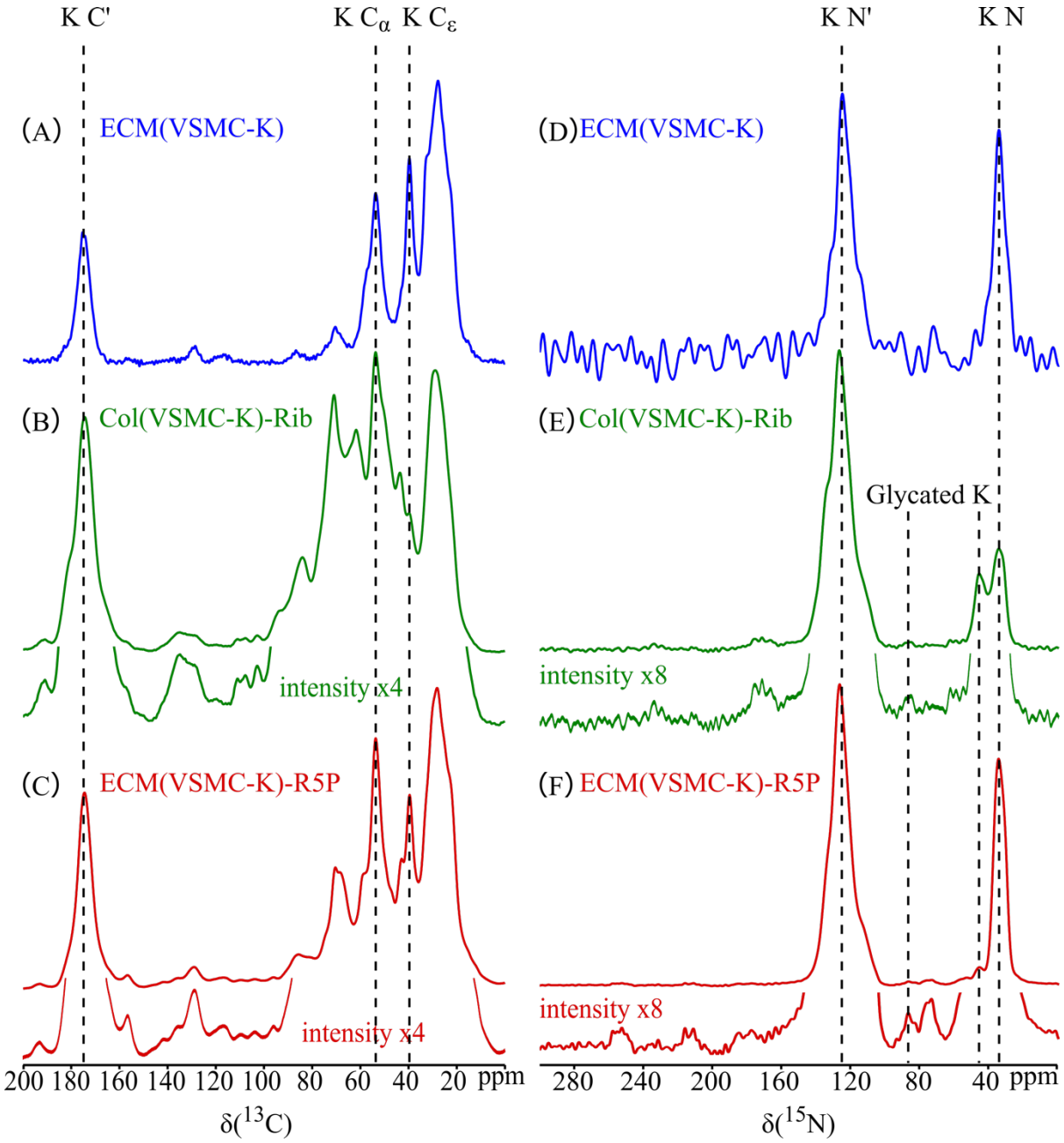


Figure 8-5 1D ^{13}C (the left column) and ^{15}N (the right column) CPMAS DNP-SSNMR spectra of ECM(VSMC-K) (blue, ns = 128 for ^{13}C 1D, ns = 16 for ^{15}N 1D), Col(VSMC-K)-Rib (green, ns = 2048 for ^{13}C 1D, ns = 6144 for ^{15}N 1D) and ECM(VSMC-K)-R5P (red, ns = 1024 for ^{13}C 1D, ns = 2048 for ^{15}N 1D). Black dash lines indicate ^{13}C and ^{15}N signals from K, and ^{15}N signals of glycated K.

Previously in the Col(FSOB-K)-R5P sample, a group of Hyl-like ^{13}C special correlations are observed in its 2D PDSD spectrum (in purple rectangles in Figure 8-4 with suggested assignments shown in Figure 8-3), while the ^{13}C connectivity between them is not revealed in the corresponding PC7 spectrum. In the 2D ^{13}C-^{13}C SQ-DQ DNP-SSNMR spectrum of ECM(VSMC-K)-R5P (Figure 8-6), the ^{13}C connectivity map in this glycated K molecule is observed (signals connected by red lines in Figure 8-6): ^{13}C(30 ppm)-^{13}C(69 ppm)-^{13}C(45 ppm), ^{13}C(44 ppm)-^{13}C (75 ppm)-^{13}C(30 ppm) and ^{13}C (71 ppm)-^{13}C(55 ppm) or, ^{13}C(44 ppm)-^{13}C (75 ppm)-^{13}C(30 ppm)-^{13}C(69 ppm)-^{13}C(45 ppm), if these signals all arise from a single molecular moiety, i.e. no accidental degeneracies between signals from different molecular species, which would lead to overlap of signal correlations. However, the ^{13}C connectivity revealed by the 2D SQ-DQ DNP-SSNMR spectrum of ECM(VSMC-K)-R5P (Figure 8-6) does not match the deduced structure from the 2D ^{13}C-^{13}C PDSD SSNMR spectrum of Col(FSOB-K)-R5P: ^{13}C(30 ppm)-^{13}C (45 ppm)-N-^{13}C(53 ppm)-^{13}C(75 ppm)-^{13}C(68 ppm) (Figure 8-3); these SQ-DQ DNP-SSNMR signal chemical shifts and connectivities are more like the previously assigned Hyl, Hyl-NH-R and glycosylated Hyl (Table 8-2).

Then, a question arises: whether the special Hyl-like ^{13}C signals are from the three possible PTMs of K or glycated K and Hyl or a combination of all of these possibilities, because all these species can give ^{13}C signals at around the above-mentioned chemical shifts. Because Hyl is widely accepted as a rare residue in collagen and not expected to be observed by SSNMR without DNP, the set of Hyl-like ^{13}C signals observed by conventional SSNMR in the R5P glycated sample (in the purple rectangle observed in Figure 8-4) was proposed to arise predominantly from the sugar moieties on glycated K and glycated Hyl, such as ^{13}C(~ 69 ppm) and ^{13}C(~ 75 ppm) and the sidechain carbons of glycated K, such as ^{13}C(~ 30 ppm) and ^{13}C(~ 45 ppm) (see Figure 8-3). However, DNP-SSNMR in principle is capable of observing ^{13}C signals from the three PTMs of K, and therefore the PTMs of K are expected to contribute to the group of signals (marked in red in Figure 8-6). Therefore, the signals marked in red in the DNP-SSNMR spectrum (Figure 8-6) are assigned as a combination of three PTMs of K and glycated K and glycated Hyl and sugar adducts.

The broadness of the ^{13}C(~ 30 ppm) - ^{13}C(~ 45 ppm) correlation signal observed in the 2D ^{13}C-^{13}C DARR DNP-SSNMR spectrum in Figure 8-7 (in red rectangles) to some extent confirms the idea that there are multiple contributors to that correlation signal.

In addition, a ^{13}C-^{13}C correlation signal pair is observed at 175-30 ppm in the 2D ^{13}C-^{13}C SQ-DQ spectrum of the ECM(VSMC-K)-R5P sample in Figure 8-6, whose chemical shifts are what would be expected for K $^{13}C'$ and $^{13}C_{\beta}$ respectively; however, SQ-DQ experiments are not likely to detect non-directly bonded ^{13}C pairs, and therefore, this correlation signal pairs is not from K nor PTMs of K but from glycation products. The 175 ppm ^{13}C signal is likely to be from an amide carbon or carboxylate carbon, and such functionalities are not likely to proceed to further reactions and usually represent the end stage of glycation reactions, i.e. in this case, AGEs.

The ^{13}C-^{13}C correlation signals, 47-26 ppm and 180-34 ppm (marked in pink in Figure 8-6 and Figure 8-7) are also observed in 2D DNP-SSNMR spectra of ECM(VSMC-K), but remain unassigned, so they are not the focus of this work. The correlation between ^{13}C(87 ppm) and ^{13}C(53 ppm) seen in the 2D SQ-DQ DNP-SSNMR spectrum in Figure 8-6 is overlapped by spinning sidebands in the 2D DARR DNP-SSNMR spectrum in Figure 8-7; these signals are most likely derived from sugar moieties in glycation products.

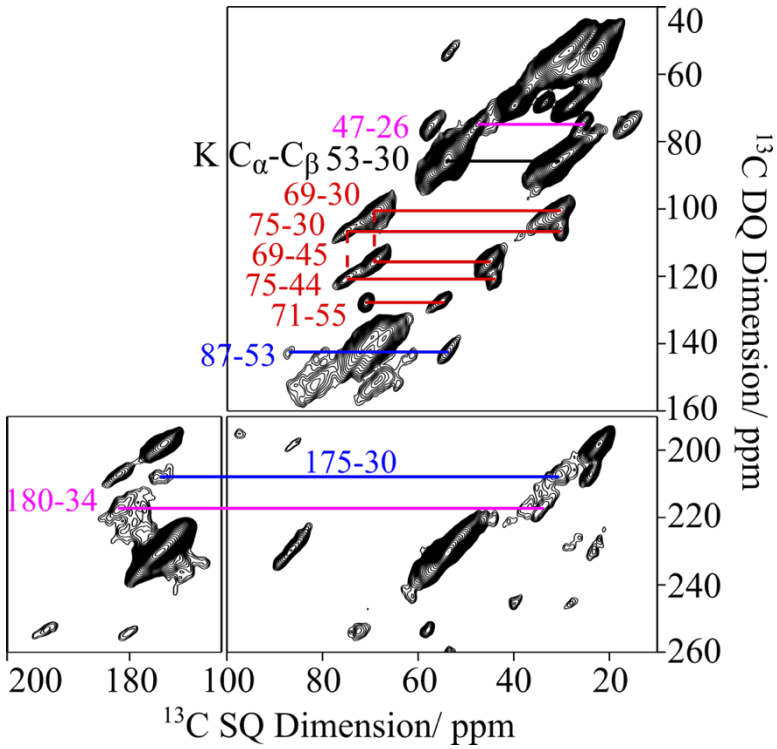


Figure 8-6 2D ^{13}C-^{13}C SQ-DQ (SPC5) DNP-SSNMR spectrum of ECM(VSMC-K)-R5P (ns = 64, 512 F1 slices).

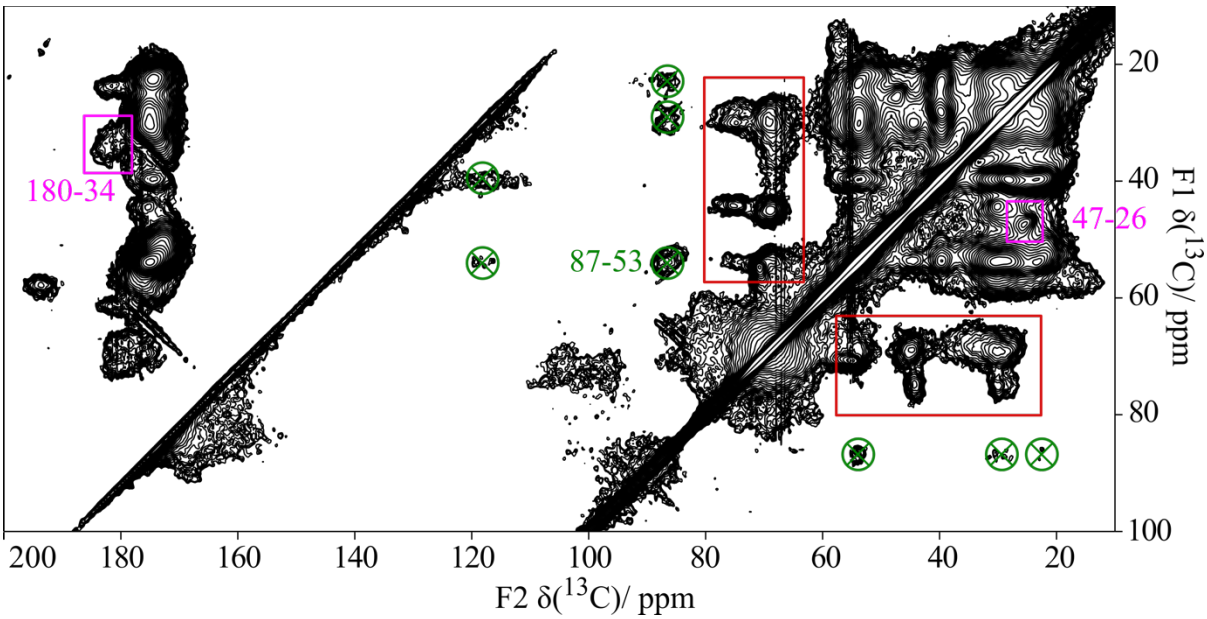


Figure 8-7 2D ^{13}C-^{13}C DARR DNP-SSNMR spectrum of ECM(VSMC-K)-R5P (ns = 64, 640 F1 slices). Signals with green marks are spinning sidebands, except the 87-53 ppm correlation signal which overlaps with signals of glycation products.

Figure 8-5 (D, E and F) also compares 1D ^{15}N CPMAS DNP-SSNMR spectra of ECM(VSMC-K) (blue, top, (D)), Col(VSMC-K)-Rib (green, middle, (E)) and ECM(VSMC-K)-R5P (red, bottom, (F)). As for ^{15}N species generated from U-$^{13}C_6$, $^{15}N_2$-K in the glycation model system, four ^{15}N signals are observed in glycated samples (Figure 8-5 (E) and (F)), 45 ppm, 57 ppm, 75 ppm and 85 ppm, in addition to the two ^{15}N signals arising from K, at 32 ppm (K ^{15}N) and 127 ppm (K ^{15}N'). ^{15}N signals with chemical shifts higher than 160 ppm (Figure 8-5 (E) and (F)) have been previously assigned to nucleotide bases [174] and therefore, we assign these signals as being from glycation species with similar chemical functionality to nucleotide bases, aromatic-like glycation products, such as argpyrimidine (structure is shown in Appendix). It is worth mentioning that sample weight of ECM(VSMC-K)-R5P is roughly 5 times more than ECM(VSMC-K)-Rib; this limited sample weight of ECM(VSMC-K)-Rib (~1-2 mg) might limit the observation of ^{15}N and ^{13}C-^{15}N correlations.

The ^{15}N signal at around 45 ppm is observed in the 1D ^{15}N CPMAS DNP-SSNMR spectra of Col(VSMC-K)-Rib (Figure 8-5 (E)) and also in the spectrum for ECM(VSMC-K)-R5P but weaker (Figure 8-5 (F)). Figure 8-8, 2D ^{13}C-^{15}N correlation DNP-SSNMR spectra, shows that this ^{15}N signal (45 ppm) correlates with two ^{13}C signals at 50 ppm and 170 ppm in spectra of both Col(VSMC-K)-Rib and ECM(VSMC-K)-R5P samples (Figure 8-8, correlation signals are connected by a black dash line), and therefore, is assigned to CML, **^{15}N**-carboxymethyl-lysine.

The ^{13}C-^{15}N correlations involving the ^{15}N signal at about 57 ppm are only observed for the ECM(VSMC-K)-R5P sample (Figure 8-8 (B), in red). This ^{15}N signal correlates with two ^{13}C signals at 47 ppm and 58 ppm. These two ^{13}C chemical shifts are very similar to P C_α and C_δ. So, these ^{13}C-^{15}N correlations might be from a P-like structure, possibly in a five-membered ring: ^{13}C-^{15}N-^{13}C.

The ^{15}N signal at 85 ppm is observed in 1D ^{15}N CPMAS DNP-SSNMR spectra of both ECM(VSMC-K)-R5P and Col(VSMC-K)-Rib samples, Figure 8-5 (E) and (F) (connected by a black dash line); however, it is only observed in the 2D ^{13}C-^{15}N correlation spectrum for ECM(VSMC-K)-R5P (Figure 8-8 (B), in red). The middle panel of the Figure 8-8 (B) shows that this ^{15}N signal (85 ppm) is in fact composed of two ^{15}N signals and which correlates with two different ^{13}C signals, ^{15}N (85 ppm)-^{13}C (40 ppm) and ^{15}N (86 ppm)-^{13}C (54 ppm), suggesting there are in fact two glycation products with an ^{15}N signal close to 85 ppm, but with

different carbon functionalities. A ^{13}C chemical shift of 54 ppm is consistent with a carbon bonded to a nitrogen and so the ^{15}N (86 ppm)-^{13}C (54 ppm) correlation signal is assigned to K ^{15}N bonded to a carbon. However, a ^{13}C chemical shift of 40 ppm is not consistent with a carbon bonded to nitrogen, so the ^{15}N (85 ppm)-^{13}C (40 ppm) correlation corresponds to a K ^{15}N and a carbon further away from that nitrogen.

A ^{15}N signal at 135 ppm correlates with the ^{13}C signal at 165 ppm, observed in 2D ^{13}C-^{15}N correlation spectra for ECM(VSMC-K)-R5P and Col(VSMC-K)-Rib in Figure 8-8, which is consistent with our previous assignment for this ^{13}C signal, -**^{15}N**H-**^{13}C**(=O)-**^{13}C**(=O)-**^{15}N**H-.

Table 8-4 The main ^{13}C-^{15}N correlations observed in 2D ^{13}C-^{15}N DNP-SSNMR spectra of Col(VSMC-K)-Rib and ECM(VSMC-K)-R5P in Figure 8-8.

^{13}C-^{15}N Correlations		**Appearance in Figure 8-8**		**Suggested Assignments or Fragments**
^{15}N /ppm	**^{13}C /ppm**	**(A)**	**(B)**	
46	50	Y	Y	CML
	170	Y	Y	
57	47	-	Y	^{13}C-^{15}N-^{13}C in a five-membered ring structure
	58	-	Y	
85	40	-	Y	
86	54	-	Y	
135	165	Y	Y	-NH-**C**(=O)-**C**(=O)-NH-

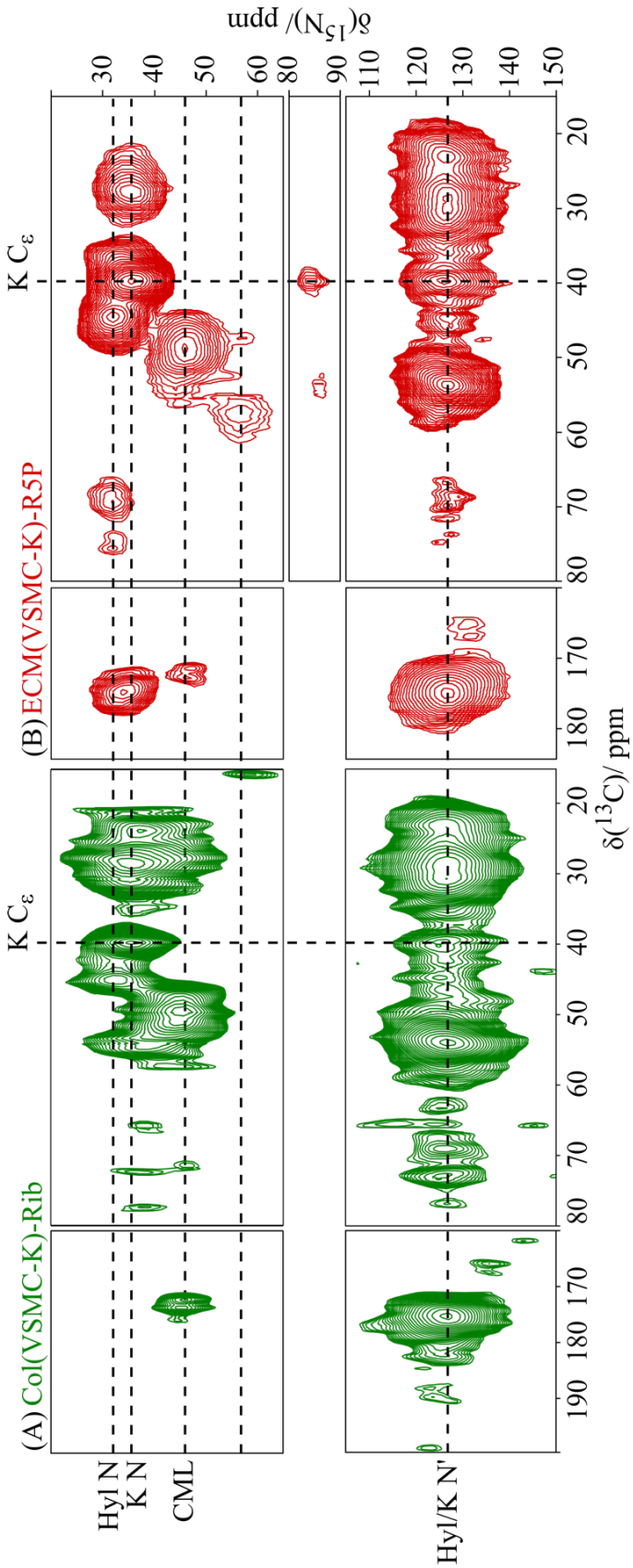


Figure 8-8 2D ^{13}C-^{15}N correlation DNP-SSNMR spectra of (A) Col(VSMC-K)-Rib (green) and (B) ECM(VSMC-K)-R5P (red). The main ^{13}C-^{15}N correlations observed are summarised in Table 8-4.

8.2 Glycation Products Derived from U-$^{13}C_6$, $^{15}N_4$-R Enriched Model Systems with U-$^{13}C_5$-Ribose-5-Phosphate and U-$^{13}C_5$-Ribose

Arginine (R) is also a target amino acid of protein glycation because it has a terminal guanidino group that can potentially participate in the glycation process. R glycation is studied in this section by examining U-$^{13}C_5$-R5P/-Rib glycated and U-$^{13}C_6$, $^{15}N_4$-R enriched in vitro ECMs using SSNMR spectroscopy, similarly to the study of K glycation in the previous section.

There are 161 R residues in the bovine collagen type I molecule (sequence is taken from UniProt), and R does not undergo complicated post-translational modifications as K does. So, in theory, all accessible guanidino groups on R are available for glycation. However, the guanidino groups on R is generally regarded to be less reactive towards glycating agents than the amine groups on K. So, U-$^{13}C_6$, $^{15}N_4$-R enriched in vitro ECMs are produced and glycated to explore whether there is any glycation reactivity between R and U-$^{13}C_5$-R5P/-Rib.

U-$^{13}C_6$, $^{15}N_4$-R enriched ECMs were prepared by in vitro cell culture where U-$^{13}C_6$, $^{15}N_4$-R is supplemented into cell culture media. Introducing U-$^{13}C_6$, $^{15}N_4$-R into cell culture can generate extra NMR signals from cell metabolism of the labelled R and thus complications in assigning NMR spectra. Cells consume U-$^{13}C_6$, $^{15}N_4$-R and metabolise to U-$^{13}C_5$, ^{15}N-P, and U-$^{13}C_5$, ^{15}N-P can be post-translational modified to U-$^{13}C_5$, ^{15}N-O in collagen. So, a combination of signals from ^{13}C, ^{15}N-labelled R, P and O can be observed in U-$^{13}C_6$, $^{15}N_4$-R enriched in vitro ECMs in addition to the signals from glycation products of interest.

Figure 8-9 compares 1D ^{13}C CPMAS SSNMR spectra of differently treated U-$^{13}C_6$, $^{15}N_4$-R enriched in vitro ECMs, including intact in vitro ECM harvested from freeze-thaw and detergent cell lysis and collagen extracted from detergent lysed in vitro ECM, to decide if different treatments would affect R conformations. R ^{13}C signals (indicated by red dash lines in Figure 8-9) do not shift much as a consequence of chymotrypsin treatment. It seems that some minor shifts are observed on R sidechain carbons, which might result from the removal of non-collagenous proteins and thus, removal of their NMR signals to make the R ^{13}C

chemical shifts more towards their chemical shifts for the triple helical conformation. The expected ^{13}C and ^{15}N chemical shifts of R in collagen are included in Table 8-5.

Table 8-5 ^{13}C and ^{15}N chemical shifts of arginine (R) observed in triple helical collagen.

AA	**δ(^{13}C)/ ppm**						**δ(^{15}N)/ ppm**		
	C'	**C_α**	**C_β**	**C_γ**	**C_δ**	**C_ζ**	**N'**	**N_ε**	**N_η**
R	174	54	28	26	42	157	120	84	72

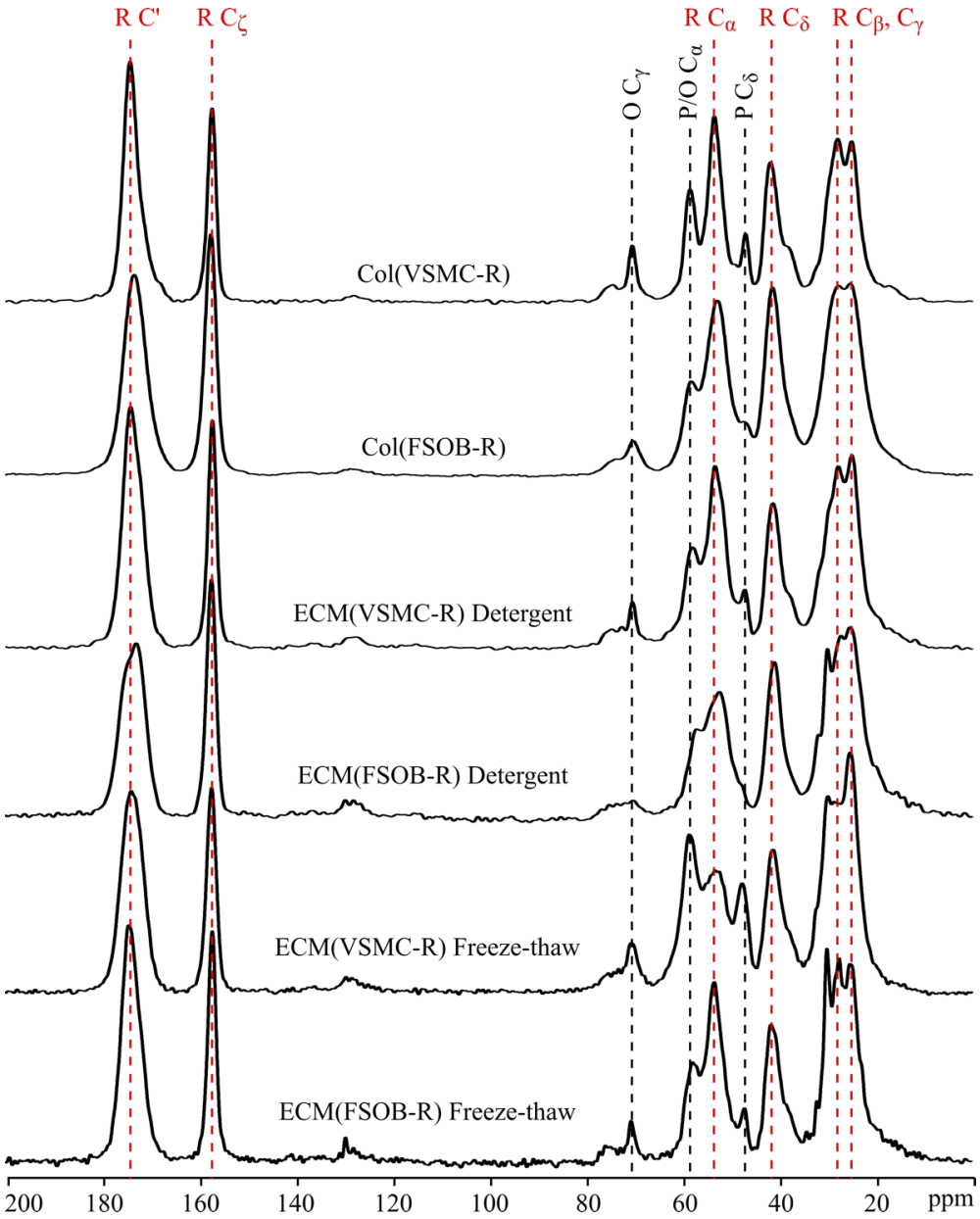


Figure 8-9 Comparison of 1D ^{13}C CPMAS SSNMR spectra of, from top to bottom, Col(VSMC-R) (U-$^{13}C_6$, $^{15}N_4$-R enriched VSMC ECM, detergent lysed, cleaned up by chymotrypsin and Tween-20, ns = 4096), Col(FSOB-R) (U-$^{13}C_6$, $^{15}N_4$-R enriched FSOB ECM, detergent lysed, cleaned up by chymotrypsin and Tween-20, ns = 4096), ECM(VSMC-R) Detergent (U-$^{13}C_6$, $^{15}N_4$-R enriched VSMC ECM, detergent lysed, ns = 4096), ECM(FSOB-R) Detergent (U-$^{13}C_6$, $^{15}N_4$-R enriched FSOB ECM, detergent lysed, ns = 31625), ECM(VSMC-R) Freeze-thaw (U-$^{13}C_6$, $^{15}N_4$-R enriched VSMC ECM, freeze-thaw lysed, ns = 4096) and ECM(FSOB-R) Freeze-thaw (U-$^{13}C_6$, $^{15}N_4$-R enriched FSOB ECM, freeze-thaw lysed, ns = 4000). Red dash lines indicate all the R ^{13}C signals, and black dash lines indicate the ^{13}C signals from O C_γ, P/O C_α (not resolved) and P C_δ.

8.2.1 Col(FSOB-R) Glycated with U-$^{13}C_5$-R5P and U-$^{13}C_5$-Rib

Glycated samples are referred to as Col(FSOB-R)-R5P or Col(FSOB-R)-Rib in the following text, and their 1D ^{13}C and ^{15}N CPMAS SSNMR spectra are shown in Figure 8-10.

Minor ^{13}C chemical shifts changes of R aliphatic carbons are observed in glycated samples compared to non-glycated samples (Figure 8-10 top), indicating R conformational changes associated with glycation. This conformational change might be related to R sidechain charge alteration due to glycation. However, the ^{13}C chemical shift changes can also arise from the production of ^{13}C enriched glycation products, giving ^{13}C signals in the aliphatic carbon region and overlap with R ^{13}C signals.

In 1D ^{15}N CPMAS SSNMR spectra of Col(FSOB-R)-R5P and Col(FSOB-R)-Rib samples (Figure 8-10 bottom), the intensities of R $^{15}N_{\varepsilon}$ and R $^{15}N_{\eta}$ decrease after glycation, when ^{15}N signal intensities are normalised on R ^{15}N', suggesting that these R N are directly involved in glycation reactions. In addition, new ^{15}N signals, weak and broad, are observed for both Col(FSOB-R)-Rib and Col(FSOB-R)-R5P samples near 100 ppm (arrow in Figure 8-10 bottom), indicating a small portion of R in those samples is glycated. Depending on the location of the R glycation site, sidechain R ^{15}N with sugar moieties attached either transitions from a "primary-amine-like" to a secondary amine or from a secondary amine to a tertiary amine; under either circumstance, the glycated R sidechain ^{15}N signal shifts are expected to result in higher ^{15}N chemical shifts than the original R ^{15}N chemical shifts.

As mentioned earlier, R can be metabolised into P by cells, and P can be post-translational modified into O in collagen. This series of activities explains the ^{13}C P and O signals observed in 2D ^{13}C-^{13}C correlation spectra (the grey solid squares and the blue dash rectangles in Figure 8-11). The ^{13}C-^{13}C correlation signals from U-$^{13}C_6$, $^{15}N_4$-R, U-$^{13}C_5$, ^{15}N-P and U-$^{13}C_5$, ^{15}N-O overlap with signals from AGEs, resulting in a group of poorly-resolved signals in the amide carbon spectral region. For example, signals in the blue dash rectangle (the amide carbon region) in Figure 8-11 (C) and (D) contain many overlapping signals, including correlations between amide carbons and aliphatic carbons in R, P and O residues as well as AGEs correlation signals 50-170 ppm (CML) and 54-176 ppm (CEL and/or MG-H and/or MODIC) observed previously.

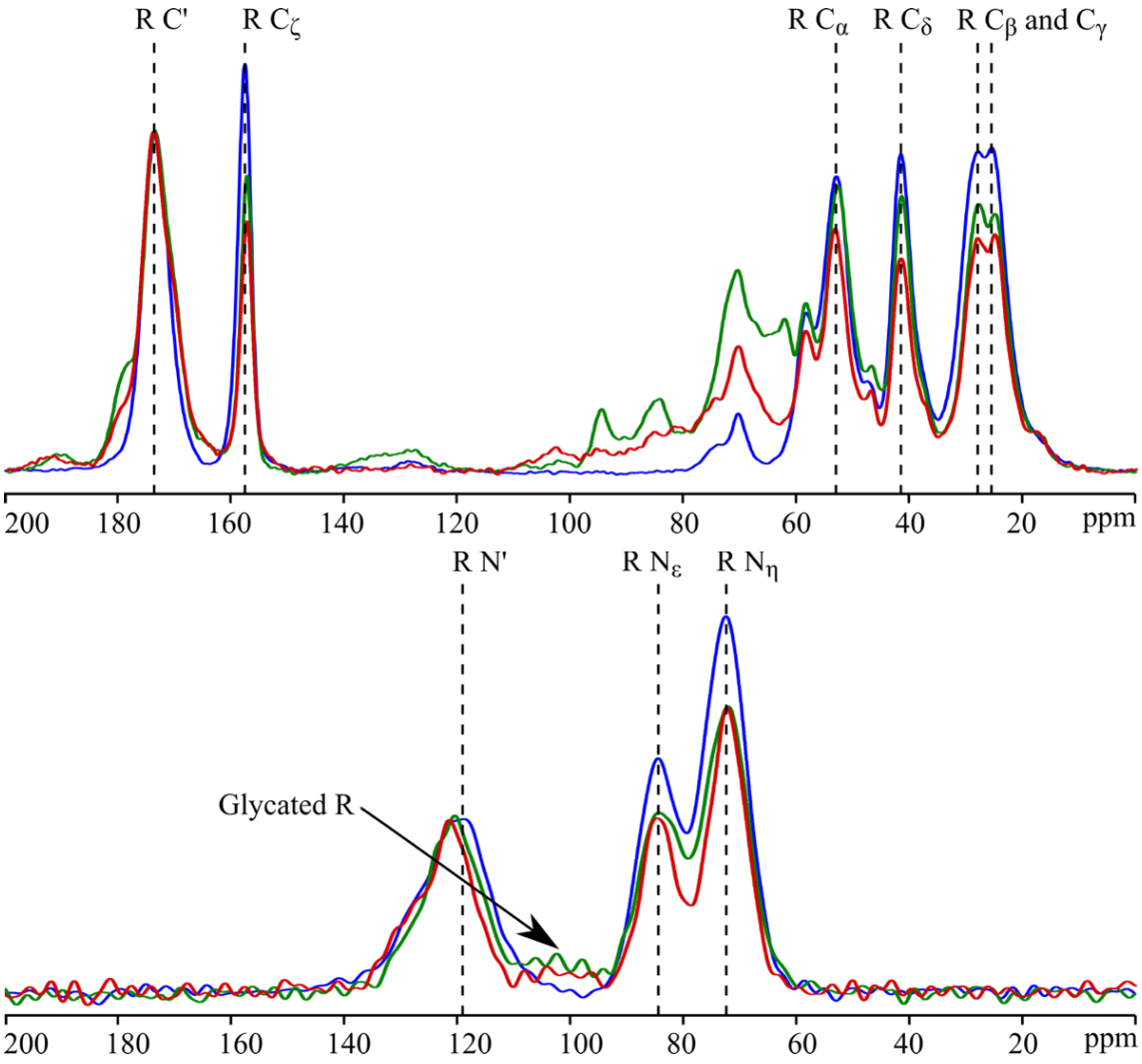


Figure 8-10 Top: 1D ^{13}C CPMAS SSNMR spectra of Col(FSOB-R) (blue, ns = 4096), U-$^{13}C_5$-Rib glycated (green, ns = 4096) and U-$^{13}C_5$-R5P glycated (red, ns = 4096) Col(FSOB-R). Intensities are normalised on R ^{13}C' signal. Bottom: 1D ^{15}N CPMAS SSNMR spectra of Col(FSOB-R) (blue, ns = 10240), U-$^{13}C_5$-Rib glycated (green, ns = 10240) and U-$^{13}C_5$-R5P glycated (red, ns = 10240) Col(FSOB-R). Intensities are normalised on R ^{15}N' signal.

However, many glycation product signals are observed and can be identified in 2D ^{13}C-^{13}C correlation spectra, PC7 and PDSD, in Figure 8-11, and ^{13}C chemical shifts of correlated signals and their suggested assignments are summarised in Table 8-6, excluding signals in the sugar and the amide carbon spectral regions. Many of the glycation products assigned here have already been observed and assigned in previously described model systems, so will not be discussed further here.

^{13}C-^{13}C correlation signals are observed at 54-75 ppm and 28-75 ppm in the 2D PDSD spectrum for Col(FSOB-R)-Rib in Figure 8-11 D. These two pairs of correlation signals share similar ^{13}C chemical shifts with previously observed Hyl-like correlations in the 2D PDSD spectrum for Col(FSOB-K)-R5P in Figure 8-4 C. The fact that the molecule/moiety giving these ^{13}C-^{13}C correlations is still ^{13}C-enriched without ^{13}C-K means that the molecule/moiety those two correlation pairs are from ^{13}C-Rib/-R5P.

Correlation signals that have not been observed in non-isotope enriched model systems are of primary interest and are most likely from glycated R (R-sugar) (chemical shifts of these signals are in bold in Table 8-6). For example, the ^{13}C PDSD correlation signal at 85-175.5 ppm seen in Figure 8-11 (D) is likely to derive from R $^{13}C_{\zeta}$ correlating with a hydroxylated carbon in the sugar adduct (R-N-$\mathbf{^{13}C_{\zeta}}$-N-$\mathbf{^{13}C_{sugar}}$(OH)), further substantiating that Rib and R5P, or by-products of their glycation with K, react with R.

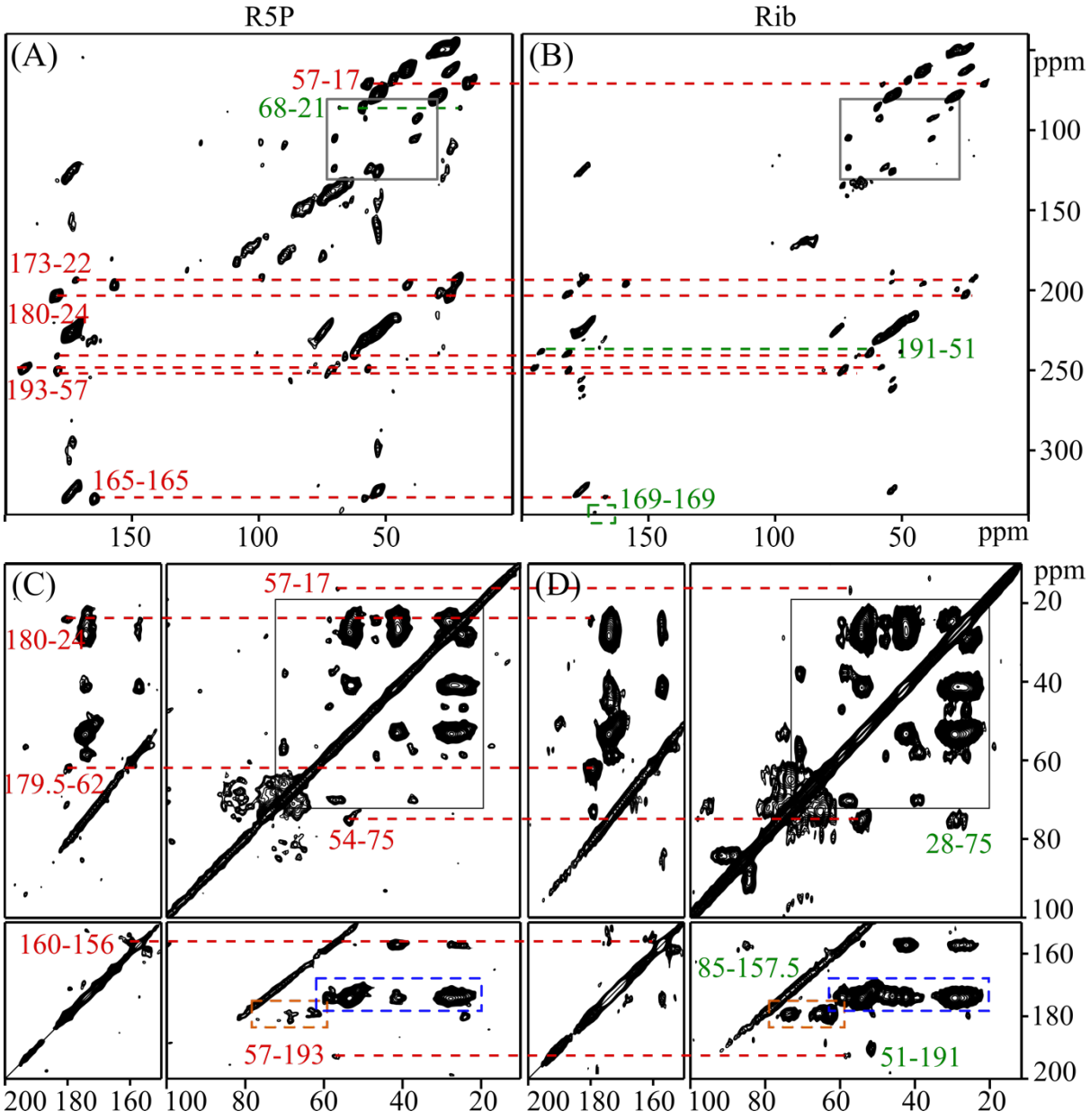


Figure 8-11 2D ^{13}C-^{13}C PC7 (top row) and PDSD (bottom row) SSNMR spectra of Col(FSOB-R)-R5P (left column; A: ns = 2048, 100 F1 slices; C: ns = 560, 278 F1 slices) and -Rib (right column; B: ns = 1280, 120 F1 slices; D: ns = 1024, 151 F1 slices). Common correlation pairs are linked by red dash lines across spectra; correlation signals only observed in one spectrum are in green. Signals included in the orange dash boxes are likely from glycation intermediates with aldehyde groups. Grey solid boxes include signals from P and O aliphatic carbons, and the blue dash rectangles include overlapping signals in amide and carboxylate region.

Table 8-6 A summary of ^{13}C-^{13}C correlation signals observed in 2D ^{13}C-^{13}C correlation spectra (PC7 and PDSD) of Col(FSOB-R)-R5P and Col(FSOB-R)-Rib in Figure 8-11. Chemical shifts in orange correspond to signals in orange dash rectangle. Chemical shifts in bold are signals that have not been observed in non-isotope-enriched model systems and are discussed in the text.

Cross Peak Correlations		**R5P**		**Rib**		**Suggested Assignments or Fragments**
Signal 1 /ppm	**Signal 2 /ppm**	**PC7**	**PDSD (20 ms)**	**PC7**	**PDSD (20 ms)**	
17	57	Yes	Yes	Yes	Yes	CEL and/or MG-H and/or MODIC
22	173	Yes	-	Yes	-	N-Acetyl-Lys
24	180	Yes	Yes	Yes	Yes	N-Acetyl-Hyl
51	191	-	-	Yes	Yes	
57	193	Yes	Yes	Yes	Yes	Norpronyl-Lys
165	165	Yes	-	Yes	-	-NH-**C**(=O)-**C**(=O)-NH-
169	169	-	-	Yes	-	
21	**68**	Yes	-	-	-	
28	**75**	-	-	-	Yes	R-sugar
54	**75**	-	Yes	-	Yes	R-sugar
57	**71**	-	Yes	-	Yes	R-sugar or sugar
85	**157.5**	-	-	Yes	Yes	R-sugar
156	**160**	-	Yes	-	Yes	
51	179	-	-	-	Yes	Glycation intermediates
62	179.5	Yes	Yes	Yes	Yes	
65	179		-	-	Yes	
69	181	-	Yes	-	-	
72	179	Yes		Yes		
73.5	179	-	-	-	Yes	

8.3 Glycation Studies in U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P Enriched Model Systems with Ribose-5-Phosphate: Conformation and Integrity of Collagen Triple Helix

G and P/O are the most abundant amino acids in collagen and are critical to collagen triple helix structure and stability. Collagen triple helix is stabilised by interchain hydrogen bonding between G N-H group and X O=C group along the backbone and the special stereo stabilising effects of P and O (reviewed in section 2.1.2). G and P/O in collagen cannot directly participate in the glycation reaction as they do not have available amino groups; therefore, their chemical structures are not affected by glycation. However, their NMR chemical shifts can be affected if glycation has an impact on the conformation of the collagen triple helix which would be observed as changes in ^{13}C chemical shifts of G and P/O, measured here in 1D ^{13}C CPMAS SSNMR spectra; if glycation disrupts the collagen triple helix, the interchain hydrogen-bonding would be affected, which leads to changes of ^{1}H chemical shifts of the hydrogen-bonded proton in G N-^{1}H groups, measured here in 2D ^{1}H-^{13}C HETCOR SSNMR spectra. Due to the special roles they play in collagen structure stability and the sensitivity of ^{13}C chemical shifts of G and P/O carbons to the molecular secondary structure, their NMR chemical shifts are used here to probe the effects of glycation on collagen and collagen dominated ECMs on collagen triple helix conformation and integrity.

In vitro ECM with U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P and U-$^{13}C_5$, ^{15}N-O enrichment is produced by cell culture with addition of the appropriately labelled G and P to the cell cultural media. Figure 8-12 compares 1D ^{13}C CPMAS SSNMR spectra of freeze-thaw and detergent lysed U-$^{13}C_2$, ^{15}N-G, U-$^{13}C_5$, ^{15}N-P and U-$^{13}C_5$, ^{15}N-O enriched in vitro ECMs, as well as collagen extracted from detergent lysed U-$^{13}C_2$, ^{15}N-G, U-$^{13}C_5$, ^{15}N-P and U-$^{13}C_5$, ^{15}N-O enriched in vitro ECMs to determine if G and P/O conformation changes in response to different treatments. ^{13}C chemical shifts of all G, P and O aliphatic carbon signals are consistent across spectra, as expected (indicated by black dash lines in Figure 8-12), except P $^{13}C_\alpha$ and O $^{13}C_\alpha$ are not resolved. Relative ^{13}C signal intensity differences between G, P and O ^{13}C in different spectra are likely to be resulted from different time points of feeding U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P into cell culture in relative to cell growth, utilisation rates of G and P and collagen biosynthesis.

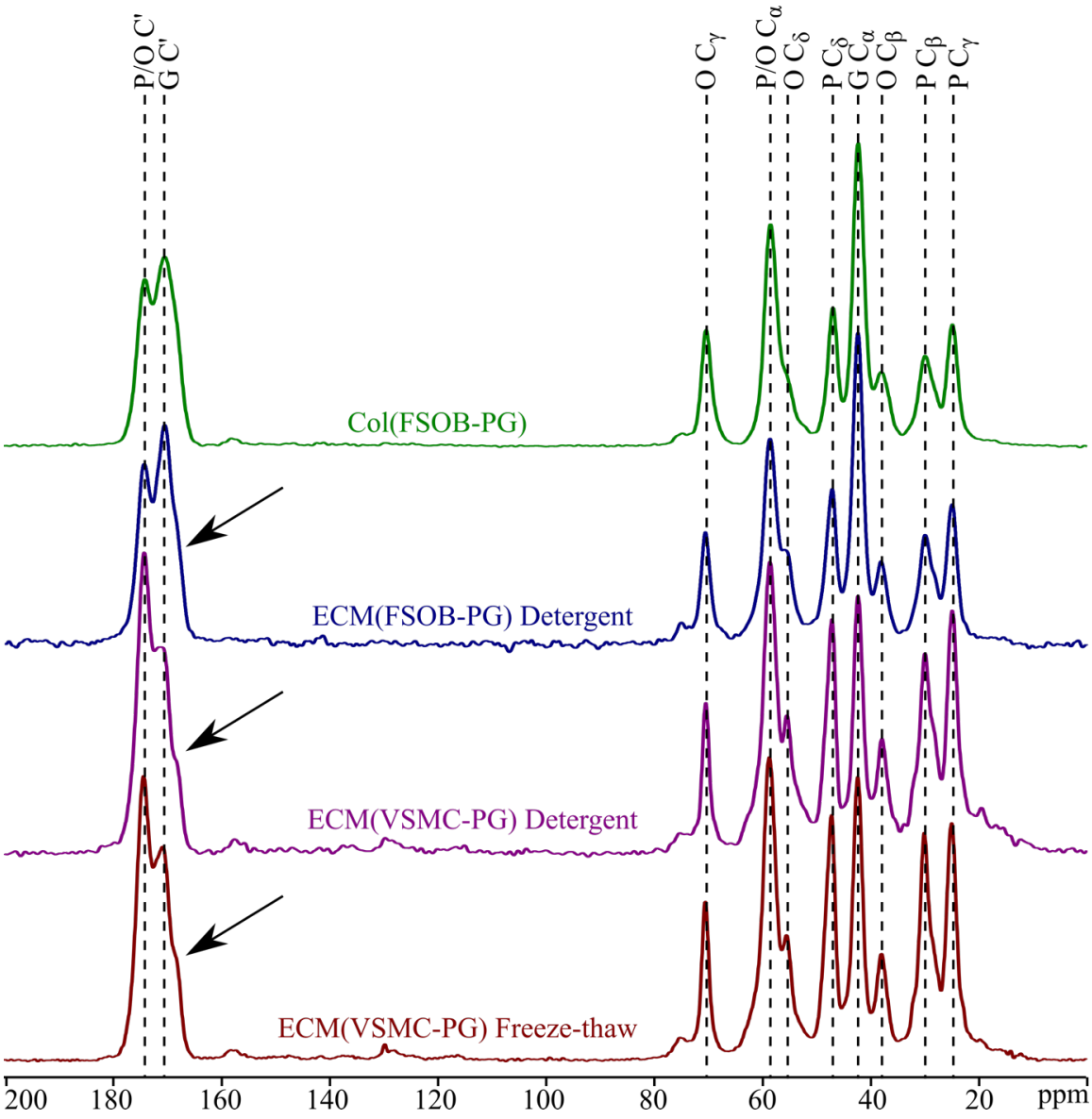


Figure 8-12 A stacking of 1D ^{13}C CPMAS SSNMR spectra of U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P labelled, differently treated ECMs, from top to bottom, Col(FSOB-PG) (green, U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P/-O enriched FSOB ECM, detergent lysed, cleaned up by chymotrypsin and Tween-20, ns = 2048), ECM(FSOB-PG) Detergent (navy, U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P/-O enriched FSOB ECM, detergent lysed, ns = 2048), ECM(VSMC-PG) Detergent (purple, U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P/-O enriched VSMC ECM, detergent lysed, ns = 4096) and ECM(VSMC-PG) Freeze-thaw (brown, U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P/-O enriched VSMC ECM, freeze-thaw lysed, ns = 4096). Black dash lines indicate the ^{13}C chemical shifts of G, P and O signals. Arrows are pointing at a shoulder signal off the major G ^{13}C' signal with lower chemical shifts, originating from G in more tightly packed triple helical regions. This colour code will be applied to samples described in the sections below.

The ^{13}C chemical shifts of P and O observed in collagenous samples shown in Figure 8-12 reflect a weighted average ^{13}C chemical shifts of P and O in different environments. ^{13}C chemical shifts of G and P/O are sensitive to local structure which is likely to be different at different points along the collagen triple helix, their neighbouring residues (see Table 8-7), as well as effects of glycation.

Table 8-7 ^{13}C chemical shifts (in ppm) of G and P/O in collagen and collagen-like model peptides. ^{13}C chemical shifts of P and O can be affected by their locations and neighbouring residues in the G-X-Y triplets and the endo or exo conformation they adopt. Chemical shifts of collagen-like triple helical model peptides were measured by Dr. Wing Ying Chow and included in her PhD thesis [135], and some are published in [175][176].

Residue Type	**δ($^{13}C'$)**	**δ($^{13}C_\alpha$)**	**δ($^{13}C_\beta$)**	**δ($^{13}C_\gamma$)**	**δ($^{13}C_\delta$)**
G in Collagen	170	42	-	-	-
P in Collagen	174	58	30	25	47
O in Collagen			38	70	55
(G**P**O)$_{11}$	171.5	58.2	27.9	25.3	46.8
(GP**O**)$_{11}$	172	58.3	38.2	70.4	55.3
(GP**P**)$_{11}$	173.7	58.9	29.5	25.5	47.5
(GPO)$_5$G**P**P(GPO)$_5$	171.4	58.1	28.2	24.8	46.7
(GPO)$_5$GP**P**(GPO)$_5$	173.9	58.9	29.5	25.9	47.6

8.3.1 ^{13}C Chemical Shifts of G and P/O

Glycated Col(FSOB-PG) and ECM(VSMC-PG) with R5P are prepared and studied by SSNMR to assess if there is any conformational changes of collagen triple helix in response to the glycation reaction. Figure 8-13 shows the 1D ^{13}C CPMAS SSNMR spectra of two pairs of comparisons between glycated and non-glycated collagenous samples with ^{13}C, ^{15}N-enriched G and P/O, Col(FSOB-PG) versus Col(FSOB-PG)-R5P^N (Figure 8-13 (A) green/bottom spectrum and black/top spectrum; nomenclature explained in figure caption) and ECM(VSMC-PG) versus ECM(VSMC-PG)-R5P (Figure 8-13 (B) purple/bottom spectrum and black/top spectrum).

After glycation, ^{13}C signals of G and P/O do not shift, indicating no significant changes in collagen secondary structure.

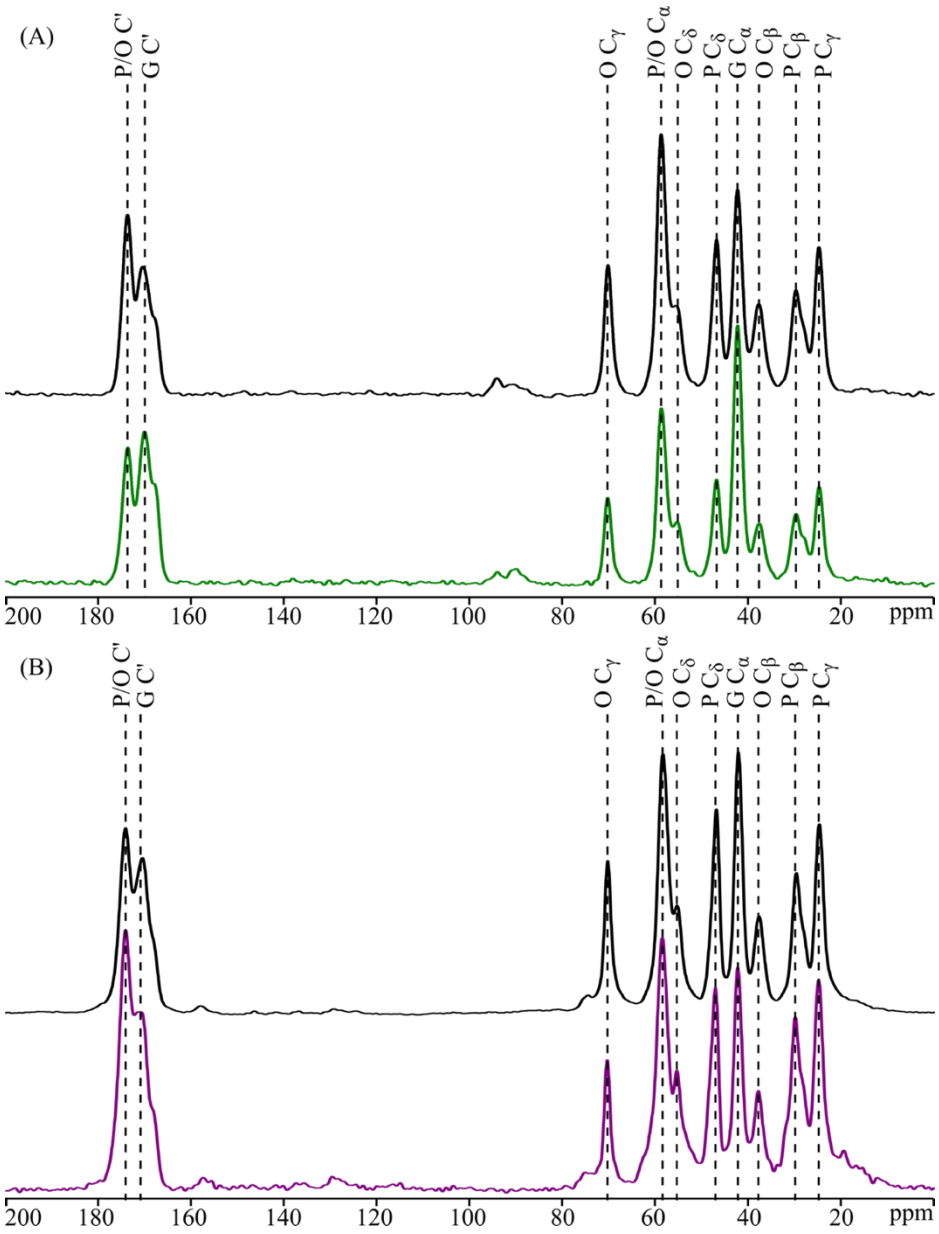


Figure 8-13 ^{13}C CPMAS SSNMR spectra of (A) Col(FSOB-PG) before (green, U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P/-O enriched FSOB ECM, detergent lysed, cleaned up by chymotrypsin and Tween-20, ns = 1024) and after (black, ns = 1024) R5P (non-isotope-enriched) glycation. Those two spectra are acquired on a 700 MHz SSNMR spectrometer at a MAS rate of 14 kHz. Both samples were prepared by Ieva Goldberga. (B) ECM(VSMC-PG) before (purple, U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P/-O enriched VSMC ECM, detergent lysed, ns = 4096) and after (black, ns = 8192) U-$^{13}C_5$-R5P glycation. Black dash lines indicate the ^{13}C chemical shifts of G, P and O signals.

8.3.2 Collagen Interchain Hydrogen Bonding

A second SSNMR measurement was trialled on glycated versus non-glycated U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P and -O enriched ECM samples to examine the effect of glycation on intrahelix hydrogen bond strength. The 2D ^{1}H-^{13}C correlation experiment gives signal correlations between ^{1}H and ^{13}C nuclei close in space and generate 2D spectra in which one dimension contains a ^{1}H spectrum and the other, a ^{13}C spectrum (more details about this SSNMR experiment is explained in section 4.6.2).

In the collagen triple helix, every G N'-H group proton is hydrogen-bonded to the carbon in the O=C' group of the residue at X position (N'-H (G)···O=C' (X)) on a neighbouring chain to stabilise the triple helix. From a SSNMR perspective, this hydrogen-bonded ^{1}H in the N'-H group should give a distinctive ^{1}H-$^{13}C_{\alpha}$(G) cross peak in the 2D ^{1}H-^{13}C HETCOR spectrum. The ^{1}H chemical shift of the ^{1}H-$^{13}C_{\alpha}$(G) signal depends on the strength of hydrogen-bonding: the longer/weaker the hydrogen-bond, the lower the ^{1}H chemical shift. If the strength of this hydrogen-bond changes with glycation, the implication is that glycation has affected the triple helix stability/integrity. We aim here to detect any such changes in G N-^{1}H ^{1}H chemical shift through correlation between the G C_{α} ^{13}C signal and the G N-^{1}H signal in the ^{1}H dimension of the 2D ^{1}H-^{13}C correlation spectra we record.

2D ^{1}H-^{13}C HETCOR spectra of Col(FSOB-PG) before and after $R5P^{N}$ glycation are presented in Figure 8-14. The G C_{α} ^{13}C signal correlates with its directly bonded ^{1}H giving a correlation signal at 3.9 ppm on the ^{1}H dimension and 42.3 ppm on the ^{13}C dimension; in the meantime, the G C_{α} ^{13}C signal correlates with ^{1}H at higher chemical shift, 8.6 ppm from the G amide ^{1}H. After glycation, the chemical shift of this G amide ^{1}H does not appear to change, suggesting no significant change on the H-bonding of this amide ^{1}H and thus glycation does not appear to affect the intrahelix hydrogen-bonding and stability of the helix from this bonding.

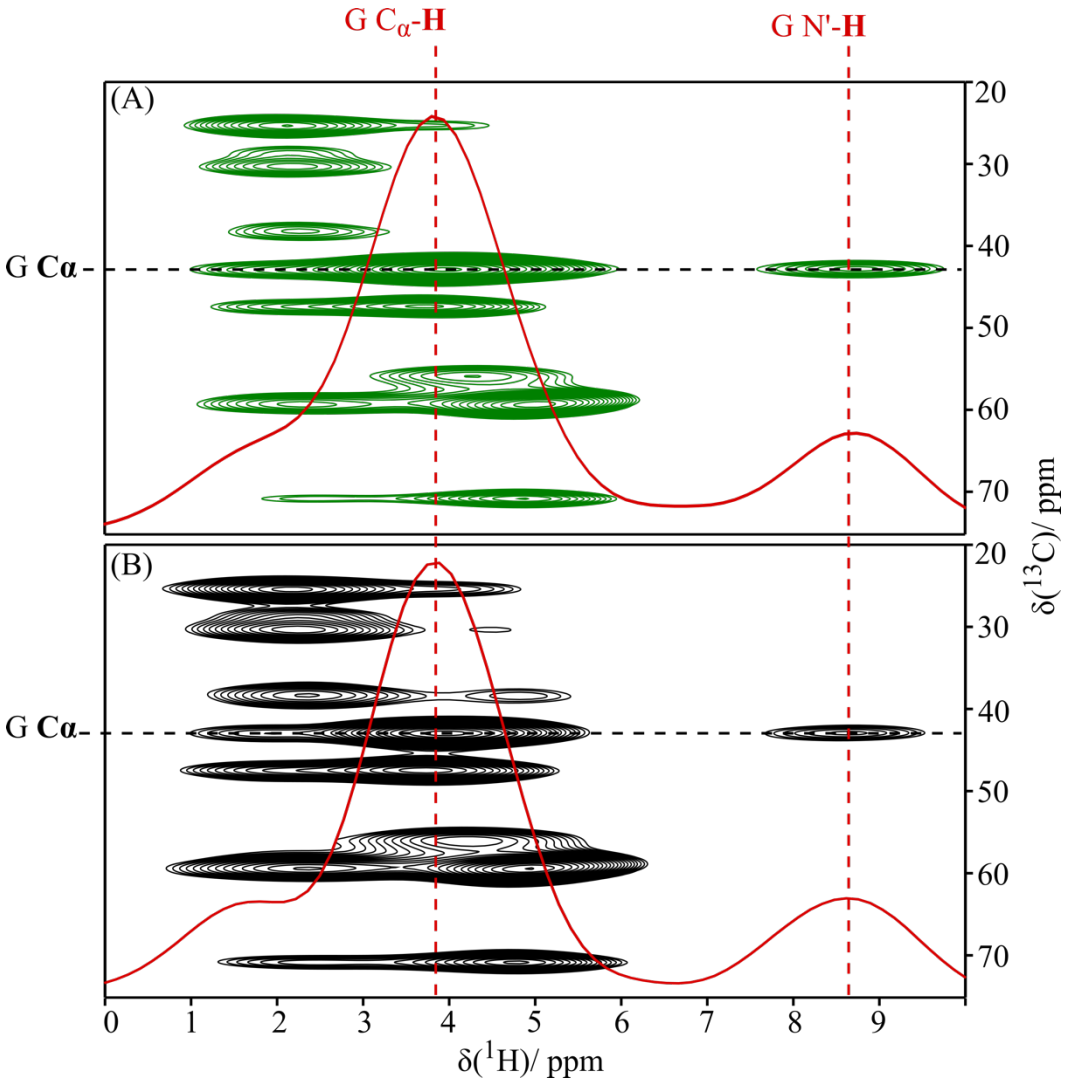


Figure 8-14 2D ^{1}H-^{13}C HETCOR spectra of (A) Col(FSOB-PG) (green, U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P/-O enriched FSOB ECM, detergent lysed, cleaned up by chymotrypsin and Tween-20, ns = 256, 50 F1 slices) (B) Col(FSOB-PG)-R5P^{N} (black, U-$^{13}C_2$, ^{15}N-G and U-$^{13}C_5$, ^{15}N-P/-O enriched FSOB ECM, detergent lysed, cleaned up by chymotrypsin and Tween-20 and glycated with ribose-5-phosphate, ns = 256, 50 F1 slices). Slices extracted at G C_α shown in red, and red dash lines indicate ^{1}H chemical shifts of G Ca-^{1}H and N'-^{1}H. Both spectra are acquired on a 700 MHz SSNMR spectrometer at a MAS rate of 14 kHz. These two samples were prepared by Ieva Goldberga.

8.4 Summary

K and R directly participate in the glycation reaction, and their chemical structures are expected to change due to glycation and glycation products adding to their sidechains. We explored changes of K and R conformation with glycation via ^{13}C and ^{15}N chemical shift changes from samples in which the collagen K and R was uniformly ^{13}C, ^{15}N labelled. ^{13}C-labelled glycating agents were used in the glycation reactions in order to extract carbon and nitrogen connectivities between K and R sidechains and glycation additions to those sidechains. ^{15}N chemical shifts of K or R are expected to change significantly with glycation, as the glycation chemistry occurs at these nitrogens, and thus ^{15}N NMR spectra were used to assess the degree to which K and R each participate in glycation reactions.

DNP-SSNMR spectroscopy compensates the intrinsic sensitivity problem of conventional SSNMR spectroscopy and allows observation of signals derived from minor glycation products that are not observable by the conventional SSNMR. 2D ^{13}C-^{15}N correlation DNP-SSNMR spectra confirms the assignments of N-carboxymethyl-lysine and the oxalamide-like structure in the previous chapter.

G and P/O do not participate in the glycation reaction, but their chemical environments can change due to glycation if glycation affects more than simply the local structure of the glycated residue. We monitored any such effects here through the respective NMR ^{13}C and G N-^{1}H chemical shifts of G, P and O and concluded that glycation has little effect on any of these parameters. We thus suggest that, rather than causing gross changes to the collagen triple helix conformation, glycation predominantly affects local collagen structures, including K and R sidechain conformations. Modifications of K and R sidechains may lead to further consequences and eventually cause tissue failure. For example, charge alterations of K and R sidechains may also alter the electrostatic interactions between collagen molecules and cause misalignments of collagen molecules and therefore destabilise collagen fibrils. In addition, modifications of R involved in collagen-integrin binding might disrupt regular ECM-cell interactions, trigger abnormal cell behaviours and affect tissue function.

9 Biogenic Glycating Agents in Cellular Models

As reviewed in section 3.4.2, many carbonyl or dicarbonyl species have been used as model glycating agents in in vitro systems, such as GO and MGO, but not all the glycation products identified in vitro have been identified in in vivo tissues. The biological relevance of carbonyl species as in vitro glycating agents is needed to establish the in vivo relevance of their corresponding glycation products identified in in vitro systems.

Cell metabolism can vary under different conditions, especially under diseased conditions, and therefore, cells can produce different compositions of metabolites. For example, the polyol pathway is known to be upregulated in diabetic conditions, and potentially generates excess glycating metabolites such as fructose and fructose-6-phosphate compared with healthy metabolism [177]. In order to explore what biological relevant glycating metabolites can be generated by cells, multiple cell models mimicking different metabolic conditions are tested in this work to see if altered metabolism alters the amount or nature of glycating metabolites.

In terms of cellular metabolites, some metabolites are readily released outside cells (referred to as extracellular metabolites) while some only exit cells under certain conditions (referred to as intracellular metabolites), such as cell necrosis or impaired metabolism. This work mainly focuses on intracellular metabolites that can leak outside cells under necrosis. There is little known about the glycation consequences of cell necrosis, which is an inherent part of many diseases, such as cancer and inflammation.

Intracellular metabolites are collected for analysis by cell lysis. Collecting the cell lysate is challenging because cell metabolism status can change very quickly in response to environmental changes; for example, cell metabolism can change immediately when cells are taken out of the incubator or when cell culture media is removed or even during the lysis process. This fast metabolic response poses challenges on collecting the cell lysate that accurately reflects the cell metabolism status before the cell lysis procedure and requires the whole cell lysis process to be conducted as fast as possible. Therefore, it is of priority to find an appropriate cell lysate extraction method to best represent cell metabolism status for the study of intracellular metabolites and their concentrations.

Performing cell lysis and subsequent procedures at low temperature can potentially slow down cell metabolism, enzymatic processes and possible secondary reactions between reactive, i.e. glycating, metabolites and surrounding proteins or lipids. To begin with, cell lysate was collected by freeze-thaw lysis, which involves cells to be frozen in a -80℃ freezer overnight then thawed at room temperature. A considerate amount of protein is collected together with cellular metabolites, even after repeated centrifugation (see section 5.6 for metabolite extraction procedure). This protein content makes the cell lysate itself a glycation system with both glycating agents and glycating substrates (protein) are present and can result in inaccurate and misleading results. In a second set of experiments, cell lysate was collected from liquid-nitrogen-lysed cells, i.e. cell cultures flash frozen in liquid nitrogen, with extra steps to separate cellular proteins and metabolites (see section 5.6 for the detailed procedure), which resulted in more effective metabolites extraction: faster cell lysis and less protein content.

The glycating capability of extracted metabolites is then investigated by two experiments. The first one is to incubate cell lysate with poly-L-lysine (PLL), where cell lysate works as a glycating agent, and PLL works as a glycation substrate. AGE fluorescence was measured at different time points over the incubation period to determine if glycation occurs and therefore if cell lysate contains glycating metabolites. This experiment provides an overall estimation of glycation capability of intracellular metabolites. The second experiment involves high-performance liquid chromatography with UV absorbance detection (HPLC-UV) and mass spectroscopy (HPLC-UV-MS) to identify and quantify glycating metabolites in the cell lysate. A pre-column derivatisation of cell lysate is introduced to make reactive, i.e. glycating, metabolites UV active.

Results obtained from above-mentioned experiments are presented and discussed in the following sections.

9.1 Glycating Capacity of Cell Lysate - AGE Fluorescence

The glycation capacity of cell lysate is first measured by incubating cell lysate with poly-L-lysine (PLL), and AGE fluorescence is followed over time to study if glycation reactions happen and produce fluorescent AGEs during incubation. Positive controls, such as PLL with sugar or sugar phosphate, were also incubated to compare their glycation behaviours with the

cell lysate. The experiments are designed to estimate qualitatively the amount of glycating metabolites in the cell lysate.

It is worth noting that (1) in this cell lysate and PLL glycation system the dominating glycation substrate, i.e. PLL, only contain K but cellular proteins remaining (explain later) in the cell lysate might contain both K and R that can also undergo glycation reaction that could skew the results (2) only a small portion of AGEs are fluorescent and therefore detected in this experiment; glucosepane, the major AGE crosslink from glucose glycation, is not fluorescent. Known fluorescent AGEs are list in the table below (Table 9-1). However, they may not all be generated in cell lysate and PPL incubation, depending on the carbonyl metabolites present in cell lysate and whether the incubation condition matches their synthetic requirement.

Table 9-1 A list of fluorescent AGEs and their excitation and emission wavelengths (λ_{ex} and λ_{em}). The numbering in this table is corresponding to Table 3-3 and Table 3-4 in previous chapters. GA is glyceraldehyde; GA3P is glyceraldehyde-3-phosphate; MGO is methylglyoxal.

Fluorescent AGEs	**No.**	**Glycating Agent**	**λ_{ex} /nm**	**λ_{em} /nm**	**Ref**
K Amino Acid Sidechain Modifications					
GA-pyridine	5	GA	326	406	[178]
K Crosslinks					
Lysine-dihydropyridinium (K2P)	17	-	343	410	[179]
Lys-hydroxy-triosidine	33	GA	354	440	[111]
Vesperlysines	27	Glc	380	440	[180]
Crosslines	28	Glc	379	463	[181]
K R Crosslinks					
Pentosidine	13	Pentose Rib, Glc, Fru and ascorbate Xylose	335	385	[80][49] [105] [111]
R Amino Acid Sidechain Modifications					
Argpyrimidine	9	MGO GA or GA3P	320	380	[182] [111]
R Crosslinks					
Arg-hydroxy-triosidine	26	GA or GA3P	331	380	[111]

9.1.1 Freeze-thaw Lysis (FTL)

FTL, as the simplest and the most commonly used cell lysis method, was first surveyed for cell lysate collection and analysis.

Vascular smooth muscle cells (VSMCs, from adult bovine) were exposed to different glucose concentrations for different time lengths, aiming to alter normal metabolism and mimic hyperglycaemia. Then, the cell lysate was collected by FTL and incubated with PLL. AGE fluorescence was measured regularly over a 24-day incubation period, and fluorescence intensity build-up curves from three cell lysate samples are shown in Figure 9-1. Cell lysate samples collected by FTL are coded for brevity (Table 9-2).

Table 9-2 A list of cell lysate samples described in this section, and their relevant details. All samples were obtained from VSMCs. [1]Day counting started after cells are confluent.

Sample Code	Cell Culture Condition	Cell Lysis
CL1-1g-F	1 g/L Glc	Freeze-thaw Lysis
CL2-9g-1d-F	1-day exposure to 9 g/L Glc[1]	
CL3-9g-6d-F	6-day exposure to 9 g/L Glc[1]	
CL4-1g-F	1 g/L Glc	

Sample CL1-1g-F (blue curve in Figure 9-1) was obtained from VSMCs cultured with 1 g/L glucose in its culture media, which is the usual cell culture condition for VSMCs; sample CL2-9g-1d-F (orange curve in Figure 9-1) was obtained from VSMCs with 1-day exposure to 9 g/L glucose before cell lysis; sample CL3-9g-6d-F (grey curve in Figure 9-1) was obtained from VSMCs with 6-day exposure to 9 g/L glucose before cell lysis. All three fluorescence intensity build-up curves from the incubated cell lysate and PLL fluctuate with a similar trend (Figure 9-1). It seems that at the beginning, these glycation systems already show some AGE fluorescence. Then, this initial fluorescence disappears and later recovered, which might indicate that fluorescent AGE formation is somewhat reversible, which is unexpected. This may indicate that the signal/noise level in this fluorescence measurement is poor or that there are some unknown fluorescent glycation intermediates generating and advancing to non-fluorescent glycation products in the cell lysate and PLL incubation.

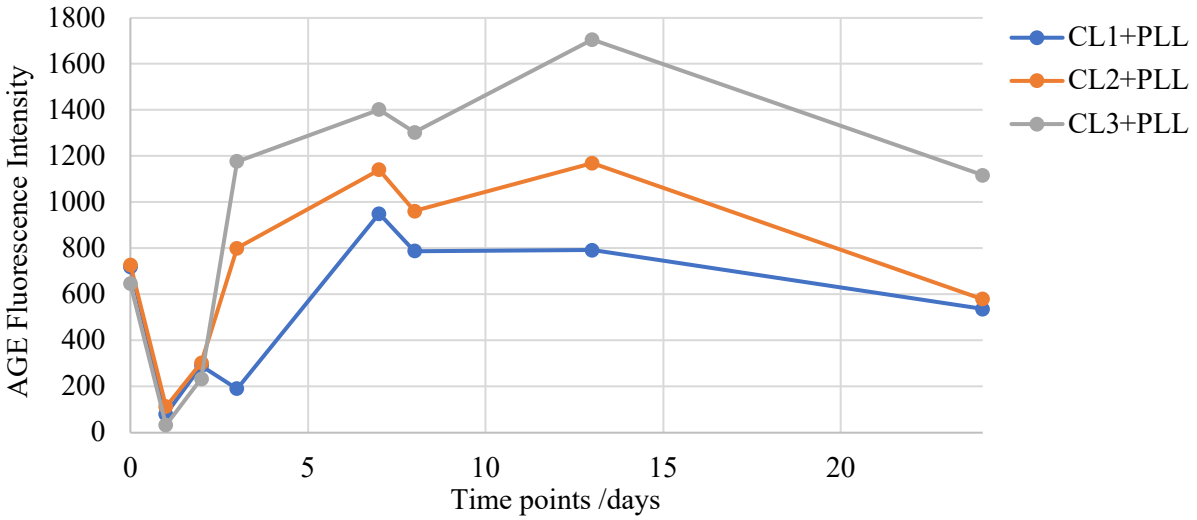


Figure 9-1 AGE fluorescence build-up curves of three different cell lysate samples and PLL over a 24-day incubation. Grey (top): CL3-9g-6d-F+PLL. Orange (middle): CL2-9g-1d-F+PLL. Blue (bottom): CL1-1g-F+PLL.

The fact that the AGE fluorescence recovers in different rates among three different CL samples and that all samples exhibit an overall increase of AGE fluorescence over the incubation period suggests the three CL samples might contain different amount of glycating metabolites. Different amounts or nature of glycating species can arise because of differences in metabolism. For example, glucose usually is metabolised via glycolysis to glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), fructose-1,6-biphosphate (F16P), glyceraldehyde-3-phosphate (GA3P), dihydroxyacetone-phosphate (DHAP), etc., which are all more reactive glycating agents than glucose; when excess glucose is provided to cells, glycolysis can be promoted and more of the reactive glucose-derived products can be produced.

After the 24-day incubation period, CL3-9g-6d-F (the grey curve in Figure 9-1) generates more fluorescent AGEs than CL2-9g-1d-F (the orange curve in Figure 9-1) and CL1-1g-F (the blue curve in Figure 9-1). This difference in the amount of accumulated fluorescent AGEs between the three samples was speculated to be a result of insufficient cell washes before cell lysis, resulting in residual glucose from the cell culture medium in cell lysate and this glucose being primarily responsible for the observed AGE fluorescence, rather than intrinsic cell metabolism differences. However, in principle, CL2-9g-1d-F and CL3-9g-6d-F have the same medium glucose concentration before cell lysis. If the insufficient washing is the sole reason for the observed glycation and AGE fluorescence, CL2-9g-1d-F and CL3-9g-6d-F should give very similar fluorescence vs time curves. To strengthen the point, glucose is known to be a very

slow glycating agent, and a minor quantity of glucose leftover from the cell culture medium is unlikely to contribute to AGEs formation in this 24-day timescale. Thus, the difference in final fluorescent AGEs accumulation suggests that cells exposed to a high glucose environment for longer periods produce more damaging metabolites as glycating agents. The final dip in all curves might result from PLL aggregation and precipitation, compromising fluorescence measurement.

Positive controls, PLL incubated with Glc, Rib and R5P, were measured in parallel to estimate the glycating capacity of cell lysate samples by comparing their AGE fluorescence intensity build-up curves (Figure 9-2). All three positive controls react significantly faster than cell lysate samples and display a higher magnitude of AGE fluorescence intensity at every time point so that they are plotted separately to show the details of the curves better. The massive difference in fluorescence intensity or AGE content of these samples is probably because the concentrations used for these sugars and sugar phosphate are too high to be comparable to the concentration of reactive metabolites in the cell lysate.

Rib+PLL (the orange curve in Figure 9-2) accumulates the most intense AGE fluorescence during the 3-day incubation, followed by R5P+PLL (the grey curve in Figure 9-2) and then Glc+PLL (the blue curve in Figure 9-2). These differences in fluorescence intensity suggest that Rib generates the most fluorescent AGEs with PLL, compared to R5P and Glc. R5P+PLL exhibits the highest AGE fluorescence intensity at the beginning of the incubation, consistent with R5P reacting more rapidly than Rib or Glc, but the AGE fluorescence does not increase as much as Rib+PLL over the incubation. The lower fluorescent AGE accumulation of R5P+PLL compared to Rib+PLL is unexpected, because R5P is widely accepted to glycate and accumulate AGEs faster than Rib. It may be that the major products of R5P+PLL at the end of the incubation period are simply not fluorescent.

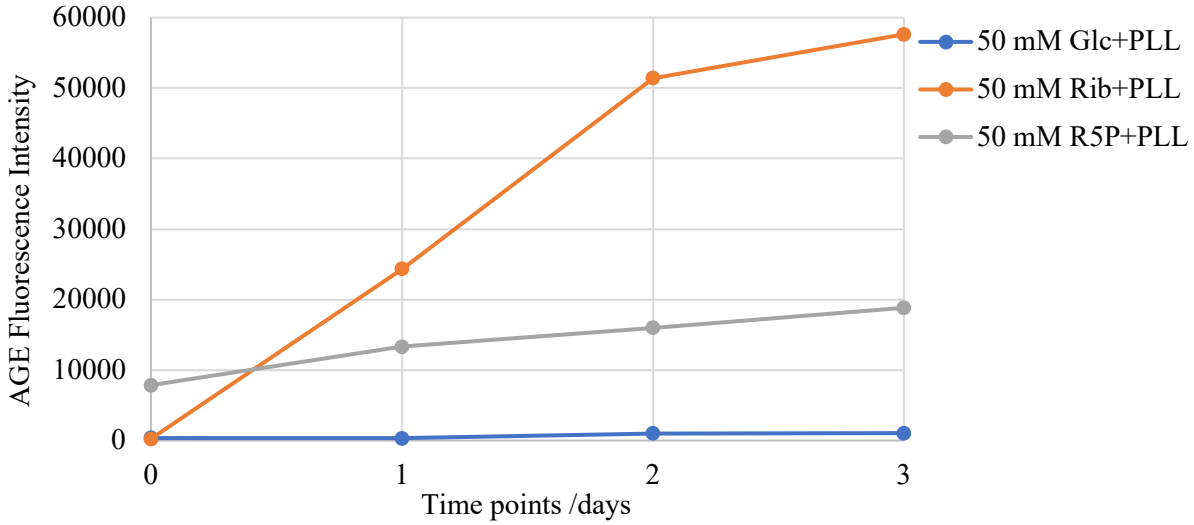


Figure 9-2 AGE fluorescence build-up curves from a 3-day incubation of 50 mM Glc, 50 mM Rib and 50 mM R5P with 50 mM PLL separately. Orange: Rib+PLL. Grey: R5P+PLL. Blue: Glc+PLL.

Protein Content in Cell Lysate

The results above and further analysis indicated that FTL resulted in a considerable amount of protein being collected together with the metabolites of interest. Consequently, the cell lysate collected by FTL contains both glycating agents, i.e. metabolites, and glycating substrate, i.e. cellular proteins, and thus results collected from above PLL incubation experiments do not reflect the glycation potential of cellular metabolites accurately. Evidence for protein in the cell lysates is presented below.

^{1}H solution-state NMR is used as a quick and crude measurement of protein presence. Protein molecules usually are much bigger than typical metabolites and have slower molecular tumbling, especially at the low temperature (5 ℃) at which the spectrum was acquired; therefore, proteins tend to give broad and poorly-resolved NMR signals while small molecules give sharp signals. Moreover, proteins have many characteristic nitrogen bonded protons (^{1}H-N) which are expected to give distinctive ^{1}H signals between 6-10 ppm. Figure 9-3 shows a ^{1}H solution-state NMR spectrum for a freeze-thaw lysed cell lysate sample, CL4-1g-F (sample code explained in Table 9-2). Broad and poorly-resolved ^{1}H signals are observed in this 6-10 ppm range, suggesting the presence of proteins in the cell lysate sample.

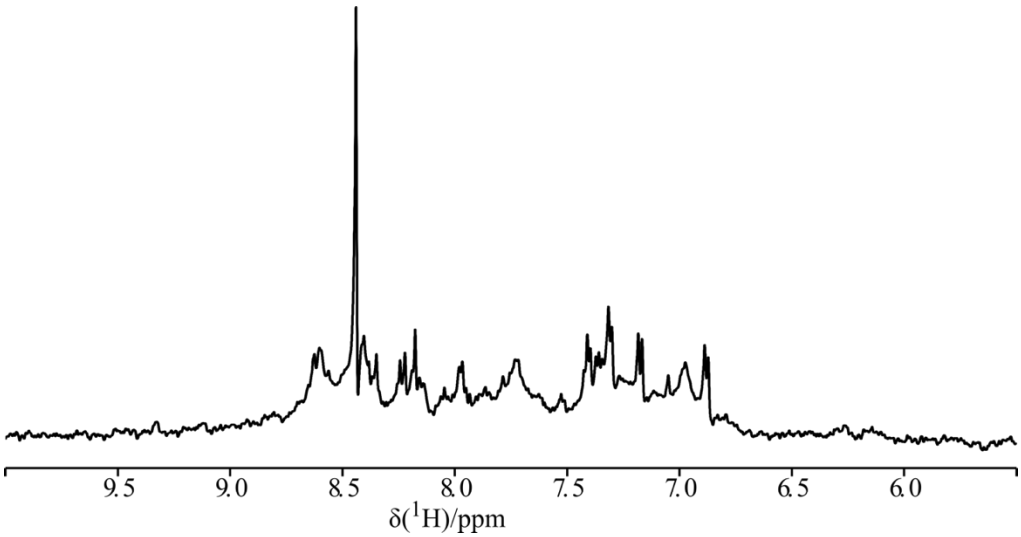


Figure 9-3 ^{1}H NMR spectrum of a freeze-thaw cell lysate sample (CL4-1g-F, ns = 1024), acquired at 5 ℃.

To confirm that proteins in FTL CL samples interfere with AGE fluorescence measurements, CL4-1g-F was incubated with either the glycating substrate, i.e. PLL, or a glycating agent, R5P. If the cell lysate has both proteins and glycating metabolites, it is expected to react with both PLL and R5P, exhibiting AGE fluorescence intensity increase upon incubation. Figure 9-4 clearly shows that cell lysate reacts with both R5P (the grey, blue and orange curves in Figure 9-4) and PLL (the yellow curves in Figure 9-4), and accumulates fluorescent AGEs over a 15-day incubation.

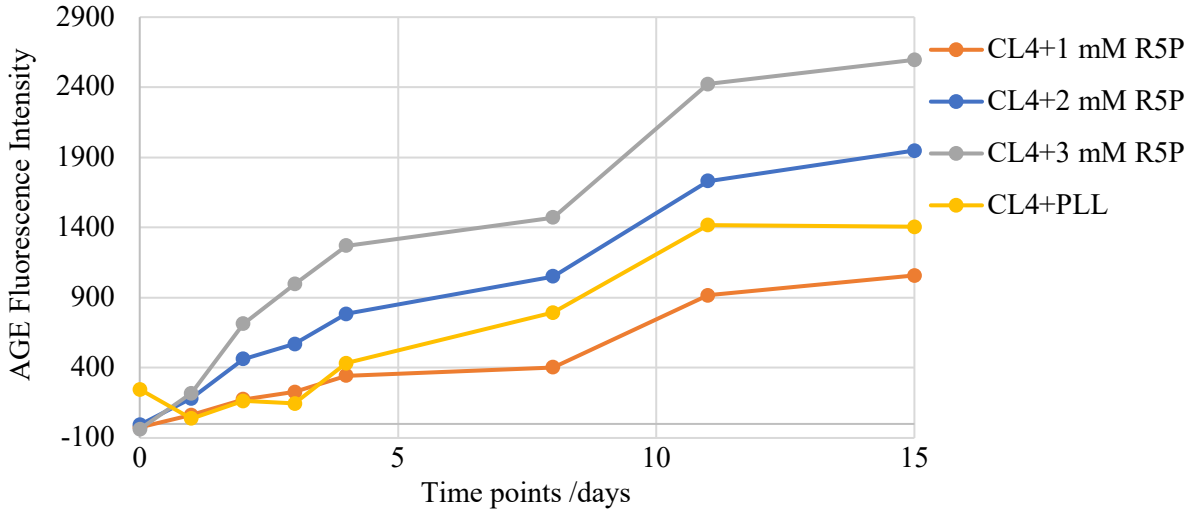


Figure 9-4 AGE fluorescence build-up curves from a 15-day incubation of cell lysate from normally cultured VSMC (1 g/L Glc) with R5P or PLL. Grey: CL4-1g-F+3 mM R5P. Blue: CL4-1g-F+2 mM R5P. Orange: CL4-1g-F+1 mM R5P. Yellow: CL4-1g-F+PLL.

Average protein content was quantified by bicinchoninic acid (BCA) assay to be 3.75 mg per flask cell (average over 4 flasks cells with each flask containing roughly 8 million cells).

Above results showed that the cell lysate collected by FTL is not only a pool of glycating agents but also a pool of glycation substrates. It can potentially react with itself and therefore, cannot result in a reliable measure representing the cell lysate glycation capacity. Under this circumstance, reactive metabolites might have already reacted with surrounding proteins before the first measurement in the PLL incubation experiment. So, a method that captures metabolites before internal reactions in the cell lysate occur is needed for this project, in which reactive metabolites should be isolated from cellular proteins as fast as possible.

9.1.2 Liquid Nitrogen Lysis (LNL)

Extraction methods to isolate metabolites from cellular proteins were researched and introduced to achieve metabolites-proteins separation and effective metabolites extraction. A protocol is evaluated in this section, involving much faster cell lysis by liquid nitrogen freezing and an additional metabolites extraction by MeOH.

In this protocol, LNL replaces the previous FTL, aiming at quenching cell metabolism and releasing intracellular metabolites immediately. Then, LNL is followed by extracting metabolites with cold 80% MeOH [183][184] and sonicating the mixture gently in ice to release as many metabolites from the cell debris as possible. Significant protein removal is achieved by the additional extraction step; ^{1}H NMR signals in the spectral region between 6-10 ppm are markedly sharpened (Figure 9-5) compared with what was observed in Figure 9-4, indicating successful removal of cellular proteins.

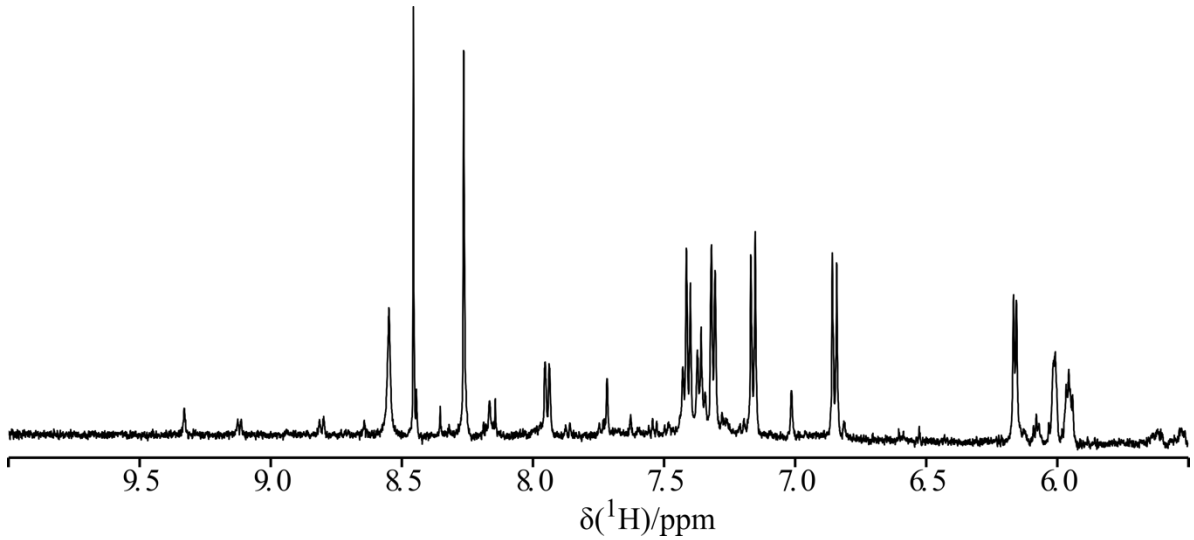


Figure 9-5 ^{1}H NMR spectrum of liquid nitrogen-lysed and MeOH extracted VSM cell lysate (ns = 1024), showing a flat baseline and a number of sharp signals in this 6-10 ppm region.

PLL incubation was repeated using cell lysate samples prepared by either FTL or LNL. Cell lysate samples discussed in this section and their sample codes are included in Table 9-3. Briefly, cells are cultured under either 1 g/L or 9 g/L glucose and lysed by FTL or LNL when they reach 90-95% confluency, roughly taking three days from seeding the cells.

Table 9-3 A list of cell lysate samples described in this section, and their relevant details. All samples were obtained from VSMCs. FTL stands for freeze-thaw lysis; LNL stands for liquid nitrogen lysis. [1]Day counting started from the day after cells are seeded.

Sample Code	Cell Culture Condition	Cell Lysis	MeOH Extraction	Sonication
CL5-9g-3d-L	3-day exposure to 9 g/L Glc[1]	LNL	Yes	Yes
CL6-9g-3d-F		FTL		
CL7-1g-L	1 g/L Glc	LNL		
CL8-1g-F		FTL		

AGE fluorescence build-up curves for different cell lysate samples and PLL incubations are displayed in Figure 9-6, as well as a curve for 50 mM (9 g/L) Glc and PLL incubation as a positive control. After a 32-day incubation, the AGE florescence intensity in the different samples is in the order: 50 mM (9 g/L) Glc (blue curve in Figure 9-6) > CL6-9g-3d-F (grey curve in Figure 9-6) > CL8-1g-F (green curve in Figure 9-6) > CL5-9g-3d-L (orange curve in Figure 9-6) > CL7-1g-L (yellow curve in Figure 9-6). This difference in the amount of

fluorescent AGEs generated by different cell lysates samples suggests that in general, LNL samples contain fewer reactive metabolites and therefore produce less fluorescent AGEs than FTL samples. It is possible that LNL quenches cell metabolism faster and more effectively in the cell lysis process than FTL and FTL puts cells under stressed or non-physiological conditions causing cells to produce reactive metabolites during the lysis process.

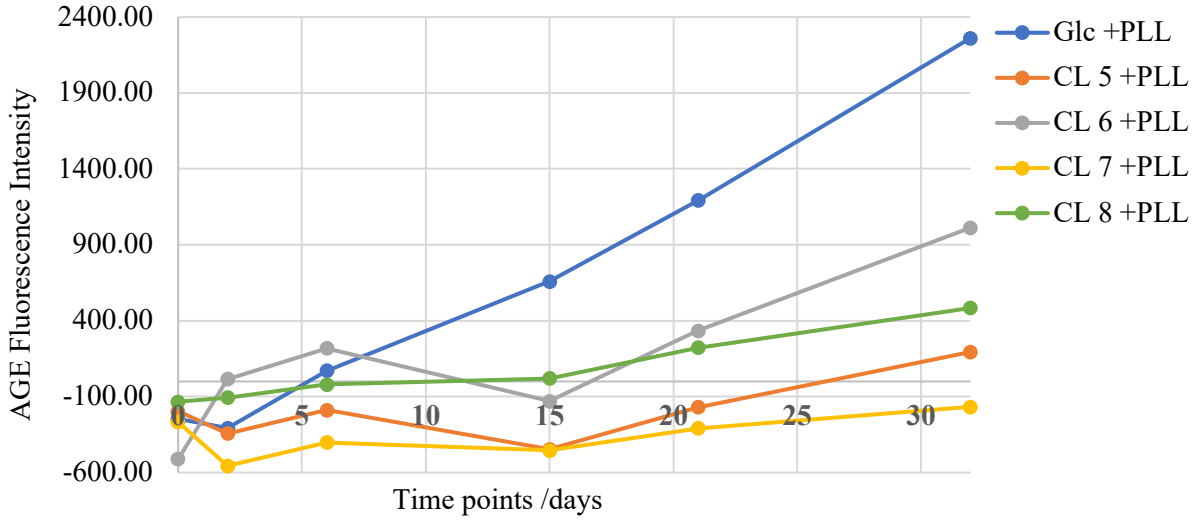


Figure 9-6 AGE fluorescence build up curves from incubating cell lysate from normally cultured VSMCs (1 g/L Glc) with Glc or PLL for 32 days. Blue: 50 mM (9 g/L) Glc+PLL. Grey: CL6-9g-3d-F+PLL. Green: CL8-1g-F+PLL. Orange: CL5-9g-3d-L+PLL. Yellow: CL7-1g-L+PLL.

9.2 Biogenic Glycating Metabolites – Identification and Quantification

A HPLC-UV-MS based protocol was established to identify and quantify what metabolites provide glycation capability to the cell lysate. Briefly, cell lysate is first derivatised by 3-amino-9-ethylcarbazole (AEC), a UV active molecule, via reductive amination (Figure 9-7) to allow the detection of reactive metabolites through UV absorption measurements. Then, metabolites with AEC adduct are separated by HPLC and verified by their retention times measured by UV absorption peaks (UV retention time) and their AEC-modified molecular weights by mass spectroscopy (MS). This method was adapted from reference [147] and optimised for the instrument used in this work.

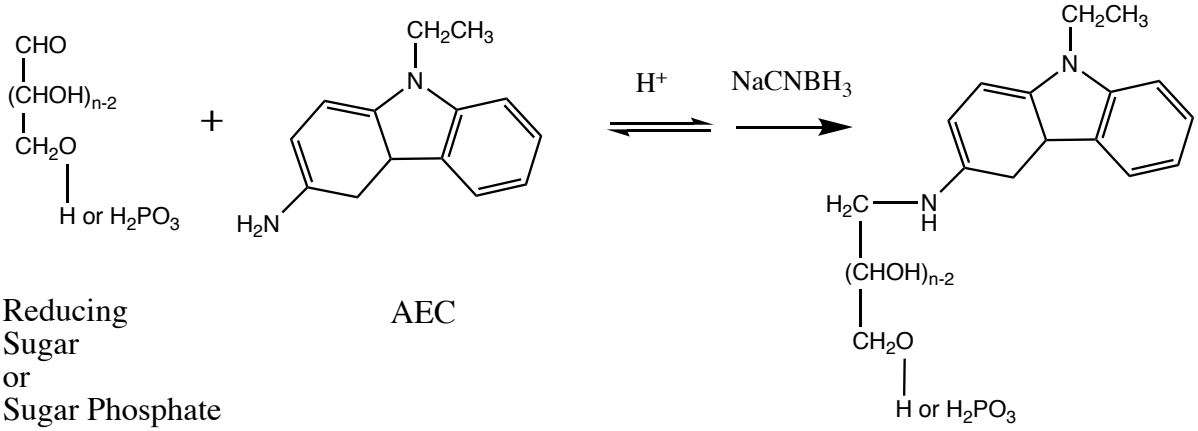


Figure 9-7 An example scheme of AEC derivatising reducing sugar or sugar phosphate. AEC can react with all carbonyl metabolites in the cell lysate, in theory, and this derivatisation scheme is similar to the initial steps of the glycation reaction.

Method Testing

Viability of this method to detect potential glycating agents was tested on a number of test glycating compounds, including glucose (Glc), fructose (Fru), ribose (Rib), glyceraldehyde (GA), fructose-6-phosphate (F6P), glucose-6-phosphate (G6P) and ribose-5-phosphate (R5P). Most of them are separated by HPLC-UV with at least 0.5 min difference in UV retention time (see Figure 9-8), and their AEC-modified molecular weights are confirmed by MS (Table 9-4) Some test compounds share the same MW, albeit with different structures such as Fru and Glc, are not distinguished by UV retention time. UV and MS peaks from excess AEC are observed in all samples.

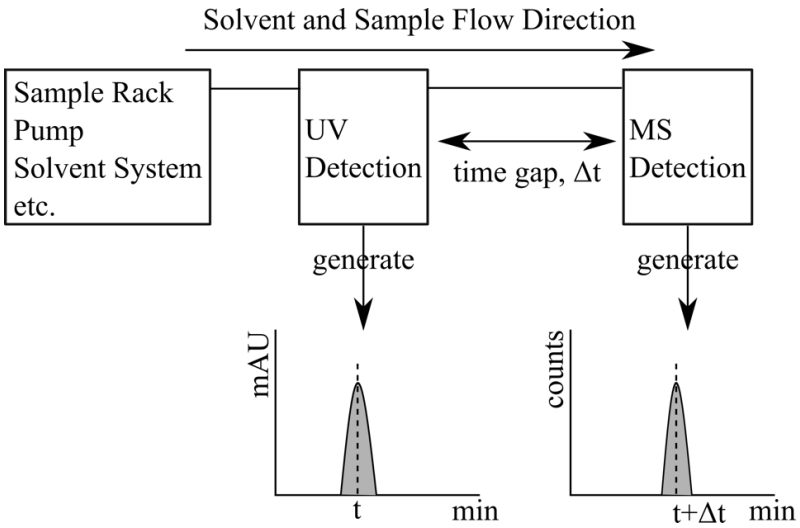


Figure 9-8 A illustration of HPLC-UV-MS work flow, where t is UV retention time and t+Δt is MS retention time. The grey area underneath a peak is integrated and used as the peak area with units of mAU*min for UV or counts*min for MS. Due to the instrumental set up shown above, a time gap exists between UV retention time and MS retention time of the same species, in this case, Δt=0.14 min

Table 9-4 UV retention times of test glycating compounds detected, including glucose (Glc), fructose (Fru), ribose (Rib), glyceraldehyde (GA), fructose-6-phosphate (F6P), glucose-6-phosphate (G6P) and ribose-5-phosphate (R5P). Fucose (Fuc) works as an internal standard for normalising UV retention time and UV absorption peak areas (integrated over time) between measurements. AEC (3-amino-9-ethylcarbazole) is the derivatisation compound to achieve UV absorption of test glycating compounds.

Compounds	**Fru/Glc**	**Rib**	**Fuc**	**AEC**	**GA**	**F6P/G6P**	**R5P**
Retention time /min	28.583	29. 287	29.783	30.383	31.933	33.443	34.830
MW-AEC-H+ /m/z	375	345	359	211	285	455	425

Then, a calibration curve between UV absorption peak areas detected at 254 nm (for the AEC moiety) and R5P concentrations was established using 2 mg/mL, 0.5 mg/mL, 0.1 mg/mL, 0.05 mg/mL and 0.025 mg/mL R5P and derivatisation time, 2 hours:

$$y = 52.298x, \mathrm{R}^2 = 0.9999 \qquad \text{(Equation 9-1)}$$

where x is R5P concentration in mg/mL and y is UV absorption peak area at 254 nm in mAU*min.

Results and Discussion

Cell lysate samples were prepared by both FTL and LNL and derivatised by AEC (sample codes in the table below, Table 9-5) to compare the differences and similarities of cell lysate from these two different cell lysis protocols by HPLC.

Table 9-5 Derivatised cell lysate samples and their details. All samples were obtained from VSMCs. FTL stands for freeze-thaw lysis; LNL stands for liquid nitrogen lysis.

Sample Code	Cell Lysis	MeOH Extraction	Sonication
LNL-AEC	LNL	Yes	No
FTL-AEC	FTL		

Many UV absorption peaks were observed in the HPLC-UV chromatographs. However, most UV absorption peaks are not intense enough to measure UV peak areas accurately nor exhibit intense enough MS peaks to assign to specific metabolites, except for AEC, Fuc-AEC and Glc/Fru-AEC peaks.

MS peaks from some metabolites of interest can however be extracted from MS spectra using their MW-AEC-H (in Table 9-4). In this way, some metabolites are identified, such as G6P, Rib and R5P. Also, MS peaks can be used to trace back to UV retention time and then UV peaks, because MS is estimated to be recorded 0.14 min after UV absorption is recorded due to the equipment setup (see Figure 9-8). For example, R5P was identified in the FTL-AEC when its MS peak was extracted between 345-345.3 m/z, and its UV retention time was traced back from MS retention time. Then, its concentration was estimated based on its UV peak area, using the calibration curve (Equation 8-1), to be 14.1 μg/mL. This concentration is converted into that roughly 8 million VSMCs (a 90-95% confluent flask) produces 0.35 μg R5P (detailed calculation is explained in Table 9-6 caption). The same attempt to estimate R5P concentration was made on the LNL-AEC sample. However, its R5P UV peak was too weak to measure its peak area. Considering its MS peak area is about 5 times less than that in the FTL sample, it is estimated that its concentration is 5 times less than 14.1 μg/mL, to be around 3 μg/mL (Table 9-6).

The HPLC-UV chromatographs in FTL-AEC samples showed more UV absorption peaks than LNL-AEC samples (sample codes are explained in Table 9-6), which indicates that FTL CL has more carbonyl metabolites than LNL CL - consistent with what was deduced in the previous section about FTL leading to cell stress and synthesis of reactive metabolites (Section 9.1).

Even though UV and MS peaks derived from other carbonyl metabolites are observed, such as Rib and G6P, it is not apparent that all reactive metabolites have the same reaction affinity towards AEC. So, for quantifying other reactive metabolites, corresponding calibration curves need to be established for each metabolite of interest.

Table 9-6 R5P concentrations and masses of two VSMC cell lysate samples, calculated from HPLC-UV-MS results. [1]This concentration is calculated from the calibration curve between the UV peak area and R5P concentration in Equation 8-1. [2]Each sample has a total volume of 100 μL, made by extracting cellular metabolites from 4 flasks of 90-95% confluent VSMCs, giving around 8 million cells each flask. (0.014 μg/μL * 100 μL / 4 flasks= 0.35 μg/flask = 0.35 μg/8M cells). [3]These two numbers are estimated from the MS peak area of R5P of the LNL-AEC sample is about five times less than that of the FTL-AEC sample.

Sample Code	**UV Peak Areas mAU*min**	**MS Peak Areas counts*min**	**R5P Concentration μg/mL**	**R5P Mass μg/8 M cells**
LNL-AEC	Too weak to measure	4399.2608	2.98[3]	0.07[3]
FTL-AEC	0.7373	20829.2433	14.10[1]	0.35[2]

9.2.1 Cancer Cell Model

A highly metastatic mouse osteosarcoma cell line, K7M2, was studied to seek more pronounced effects of changing metabolism on the amount of carbonyl metabolites cells produce. Previously, DNA and RNA have been identified in lyophilised K7M2 FTL cell lysate [174], which was never seen in other cell types used in this project. The presence of DNA and RNA in K7M2 cell lysate indicates that K7M2 is under quick and continuous proliferation, involving fast replication of their DNA (for cell proliferation) and RNA (for protein synthesis for cell growth). This difference between K7M2 and non-cancerous cell lines suggests the production of pentose and pentose phosphate might be upregulated in K7M2 to support its fast proliferation.

Method Optimisation

In previous HPLC-UV-MS experiments, we found that UV absorption peaks from glycating metabolites are not intense enough to quantify their peak areas accurately while MS peaks are relatively simple to identify. So apart from the correlation between UV peak areas detected at 254 nm and R5P concentrations (Equation 8-2 below), MS peak areas between 345-345.3 m/z and R5P concentrations are calibrated here as well (Equation 8-3 below). R5P concentrations used were 1 mg/mL, 0.5 mg/mL, 0. 25 mg/mL, 0.1 mg/mL and 0.05 mg/mL R5P, and derivatisation time was 1 hour. Also, the derivatisation reaction time was optimised on R5P to be 1 hour, rather than the previously used 2 hours.

$$y = 70.314x, \mathrm{R}^2 = 0.9948 \qquad \text{(Equation 9-2)}$$

where x is R5P concentration in mg/mL and y is UV peak area at 254 nm in mAU*min.

$$y = 1081151.72x, \mathrm{R}^2 = 0.9599 \qquad \text{(Equation 9-3)}$$

where x is R5P concentration in mg/mL and y is MS peak area between 345-345.3 m/z in counts*min.

Results and Discussion

K7M2 cell lysate was prepared by both FTL and LNL, modified by AEC and examined by HPLC-UV-MS (sample details are in Table 9-7 top). Similar to what was seen before, the peak

derived from Glc-AEC is much stronger than other metabolite-AEC peaks, except for Fuc-AEC, the internal standard, and excess AEC. Most UV peaks are weak and can only be identified when traced back from MS peaks. Therefore, although the R5P concentration in derivatised cell lysate samples, LNL-AEC and FTL-AEC, is calculated from both UV peak areas and MS peak areas using Equation 8-2 and Equation 8-3 respectively, the results from MS peak areas may be more reliable, considering UV peaks are mostly around the noise level and heavily overlapped. The R5P concentration determined by both UV peak areas using Equation 8-2 and MS peak areas using Equation 8-3 is tabulated in Table 9-7 (bottom). The R5P concentration determined in LNL CL samples is averaged between two independent (biological) repeats to be 7.05 μg/mL; while in the FTL CL sample, it is 28.8 μg/mL. These concentrations are converted to 0.235 μg and 0.96 μg per eight million cells, respectively as above (Table 9-7).

MS peak areas depend on not only the amount of metabolite-AEC present in the sample but also their responses to ionisation in the MS. So, the calibration curves between MS peak areas and R5P concentrations cannot be used for other metabolite-ACEs. Individual calibration curve needs to be established to quantify other carbonyl metabolites.

Table 9-7 Top: K7M2 CL sample code and sample details. Bottom: R5P concentrations and masses in K7M2 cell lysate were quantified using MS peak areas (in yellow shading) and UV peak areas (in blue shading) using separate calibration curves. [1]This is calculated in the same way as explained in Table 9-6; except these samples are made from 3 flasks cells. [2]This is the average between two repeats.

Sample Code	**Cell Lysis**	**MeOH Extraction**	**Sonication**
LNL-AEC-1	LNL	Yes	No
LNL-AEC-2			
FTL-AEC	FTL		

Sample Code	**MS Peak Areas counts*min**	**R5P Concentration µg/mL**		**R5P Mass µg/8 M cells**	**UV Peak Areas mAU*min**	**R5P Concentration µg/mL**		**R5P Mass µg/8 M cells**
LNL-AEC-1	12171.76	11.26	7.05[2]	0.235[1]	0.4811	6.84	5.00[2]	0.17[1]
LNL-AEC-2	3075.87	2.84			0.2214	3.15		
FTL-AEC	31132.52	28.80		0.96[1]	1.1391	16.20		0.54[1]

9.3 Summary

Cell lysate was collected to study the glycating capacity of cellular metabolites in order to estimate the glycation damage under cell necrosis in diseases and establish the relevance of glycating agents used in previous in vitro studies and in this work.

A healthy primary cell line, VSMC, and a cancer cell line, K7M2, were used to study the glycation potential of their metabolites, with focuses on establishing the relevance of R5P as glycating agents under cell necrosis conditions and estimating the R5P concentration in different cell lysates. Cell lysates from VSMC and K7M2 were collected by two different cell lysis methods and gave different results regarding their glycation capability and R5P quantities. Freeze-thaw lysis (FTL) is straightforward and sufficient for lysing cells and retrieving metabolites but insufficient in removing cellular proteins or effectively capturing snapshots truly representative of cell metabolism. However, the cell lysate collected by FTL might mimic stressed cell metabolism status because the results from this model suggested that FTL puts cells under stress during the slow freezing period and causes them to produce more glycating metabolites. The liquid-nitrogen lysis (LNL) protocol contains multiple steps but appears to be more suitable for this project. The LNL acts faster and more effectively than FTL in shutting down cell processes and thus better reflect cell metabolism status; the additional extraction step successfully eliminates cellular proteins and extracts cellular metabolites.

Glycation potential of cell lysate is measured by incubating cell lysate with PLL and following the accumulation of AGE fluorescence over the incubation period. Our results conclude that cell lysate contains glycating agents, and their glycation capacities appear to be affected by the time length of exposure to high glucose concentration during cell culture and the speed of cell lysis. Extended exposure to high media glucose concentration and slow cell lysis appears to make cells produce more glycating agents and show strong glycating power in the cell lysate. However, PLL incubation experiments are rather qualitative because of the possible interference of fluorescence measurement from protein aggregation and possible colour changes.

Glycating cell metabolites were further investigated by HPLC-UV-MS to identify and quantify specific carbonyl metabolites. R5P was quantified in both healthy and cancer cell models as a

powerful potential glycating agent. In addition, a general trend is observed that FTL cell lysate samples have more R5P than LNL cell lysate samples, and K7M2 cell lysate samples have more R5P than VSMC cell lysate samples. However, whether this trend is statistically relevant cannot be concluded from the data here. Many more independent and biological repeats of cell lysate samples are needed to work out averages and standard deviations of R5P concentrations in different cell models and cell lysis protocols before reaching a solid conclusion.

10 Conclusion and Future Work

Glycation chemistry and glycation products have been studied extensively because of their implications in the development of diseases, such as diabetes and the development of diabetic complications, eventually affecting many tissues and organs. Nevertheless, the underlying chemistry is not yet fully understood due to the complex glycation reaction routes and numerous biogenic glycating agents. So, a better understanding of glycation chemistry and identifications of biogenic glycating agents and glycation products are important to study many pathologies and offer possibilities of early diagnoses and novel therapeutics.

This project aims to study the glycation chemistry and identify glycation products, especially glycation intermediate products, in two in vitro collagenous ECM models, from vascular smooth muscle cells and osteoblasts, to mimic the effects of glycation in blood vessels and bone tissues using a combination of isotope labelling and solid-state nuclear magnetic resonance (SSNMR) spectroscopy. A combination of ribose and ribose-5-phosphate are used as glycating agents, which also models the effects of glycation by reducing sugars and reducing sugar phosphates. A number of glycation products has been identified, and structures of some glycation products have been suggested based on their NMR chemical shifts. The major glycation products identified in two types of collagenous ECM studied in this project appear to be very similar, which are lysine and arginine sidechain modification products, such as N-carboxymethyl-, N-carboxyethyl- and N-acetyl- adducts, rather than the widely-studied intermolecular crosslinks. Therefore, it is possible that the observed changes of collagen accompanying the glycation reaction are caused by the copresence of amino acid sidechain modifications and intermolecular crosslinks rather than crosslinks alone, which are in relatively low quantities. Though some SSNMR glycation product signals are still unassigned, the work has demonstrated that a huge variety of glycation products, including glycation intermediates and advanced end products, are present in glycated collagenous ECM, giving yet further insight into the complexity of glycation chemistry. Clearly, there is further work to do to identify the unknown glycation products from the as-yet unassigned NMR signals and to investigate their roles in the consequences of glycation.

In contrast to previously used techniques, such as immunohistochemistry and LC-MS, we show here that SSNMR spectroscopy can capture the complex equilibria of the glycation reaction and has the potential to examine all the glycation products present in the collagenous ECM samples. This map of all glycation products with a focus on major glycation products so far has not been achieved by other techniques, because most techniques can only measure AGE concentrations of selective AGEs and lack of methods to identify or quantify glycation intermediates. This lack of glycation product profiles also means that the chemical species lead to the observed detrimental effects of glycation cannot be determined. Overall, SSNMR spectroscopy offers insights on the structures of glycation products, relative quantities of different glycation products generated in collagenous materials and might in the future contribute to the investigation of causes of glycation damages and the development of protocols to quantify the glycation intermediate products by more sensitive techniques, such as LC-MS. However, SSNMR spectroscopy has its intrinsic disadvantages regarding low sensitivity, long experiment time as well as high dependence on isotropic enrichment. This means SSNMR is not likely to detect glycation products present in minor quantities nor non-isotopically enriched fragments in glycation products. The ongoing development of DNP-SSNMR spectroscopy overcomes the disadvantages of conventional SSNMR spectroscopy to some extent and can offer new insights into the study of the glycation chemistry.

With enhanced SSNMR sensitivity given by dynamic nuclear polarisation (DNP), DNP-SSNMR spectroscopy contributes further to glycation products identification, especially by experiments observing ^{15}N and ^{13}C-^{15}N correlations in the glycation model systems. Some ^{15}N signals derived from U-$^{13}C_6$,$^{15}N_2$-lysine modifications in glycated collagenous ECM that have not been detected by conventional SSNMR spectroscopy are detected by DNP-SSNMR spectroscopy, such as carboxymethyl-lysine, an oxalamide-like fragment and a proline-like ring fragment. However, not all ^{15}N signals observed in DNP-SSNMR spectra are assigned to specific glycation products because ^{15}N chemical shifts are not as well studied as ^{13}C chemical shifts in general and cannot be easily predicted, and thus assigned, by existing tools. Further ^{15}N signal assignments might be facilitated in the future by synthesising and studying ^{15}N chemical shifts on model compounds structurally similar to predicted glycation products.

In addition to identifying glycation products, the molecular conformation and integrity of collagen triple helix does not appear to be affected by the glycation reaction according to SSNMR data acquired in this project. The retention of an intact, largely undistorted collagen

triple helix after glycation suggests that consequences of collagen glycation are more likely to result from changes of K and R sidechain, including structural changes, charge alterations and conformational changes, and subsequent changes of collagen higher level structure, such as fibril alignments, rather than from significant disruption of collagen molecules themselves.

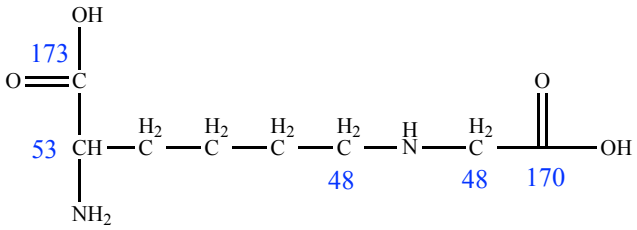
In terms of the exploration of potential biogenic glycating agents, R5P is the focus in this work as it is an essential metabolite in nucleic acid metabolism and a vigorous glycating agent. A HPLC-UV-MS based protocol was established in this thesis to semi-quantify the R5P content circulating in cells to gain insight into the amount of R5P that could be released into the ECM. The results demonstrated that the R5P level is elevated under stress conditions and in cancer cells compared with healthy cells, potentially leading to increased R5P glycation of the ECM in diseased and cancerous tissues where there is cell necrosis. The HPLC-UV-MS protocol established in this work to quantify R5P in cell lysate can be easily applied to quantify other glycating metabolites, such as glucose-6-phosphate, and may also be applied to study the glycation supernatant in this model system to identify and quantify the soluble (di-)carbonyl glycation intermediates or sugar fragments to further study glycation mechanisms and advance glycation studies.

The glycation reaction has been implicated in many diseases though its chemistry is not fully understood. Previous studies have attributed to the deleterious effects of glycation to the formation and accumulation of AGEs. However, without the investigation on the glycation intermediates, whether the glycation damages are purely caused by AGE accumulation cannot be determined. This thesis offers valuable insights on the presence of glycation intermediate products, which appears to be the major glycation products generated in the in vitro ECM and Rib/R5P glycation reaction. The employment of SSNMR spectroscopy allows revealing the structures of glycation intermediate products. The relative abundant presence of glycation intermediate products in the complex glycation equilibria suggests that the intermediate products are expected to contribute to glycation damages.

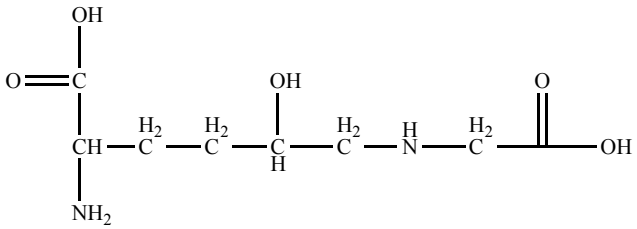
Appendix: Structures and ^{13}C NMR Chemical Shifts of AGEs

This appendix includes all the glycation products mentioned in Table 3-3 and Table 3-4, including their full names, abbreviations used in the main text, structures and their ^{13}C chemical shifts extracted from solution-state NMR spectra, if known. ^{13}C and ^{15}N SSNMR chemical shifts identified in this project are in red, and literature values are in blue.

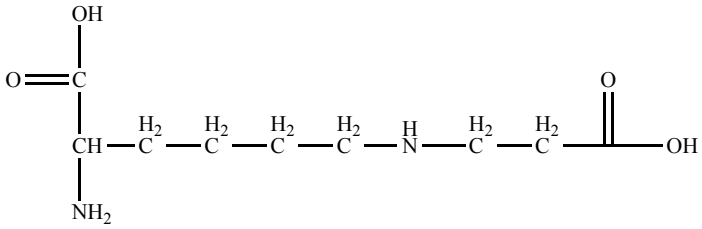
1. N-carboxymethyl-lysine (CML) and ^{13}C chemical shifts [185]



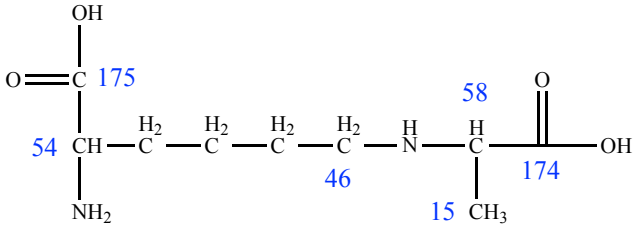
2. N-carboxymethyl-hydroxylysine (CMhL)



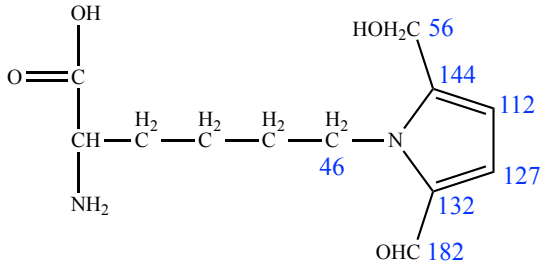
3. A: N-carboxyethyl-lysine (CEL)



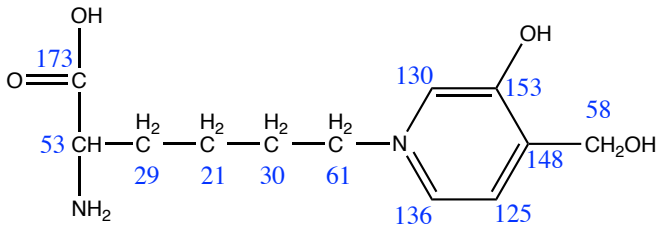
B: N-1-carboxyethyl-lysine (CEL) and ^{13}C chemical shifts [186]



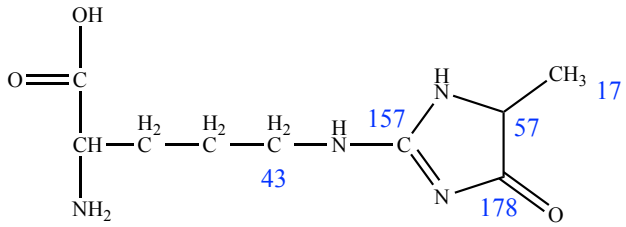
4. Pyrraline and ^{13}C chemical shifts [187]



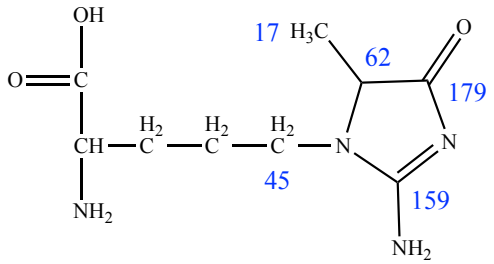
5. GA-pyridine and ^{13}C chemical shifts [100]



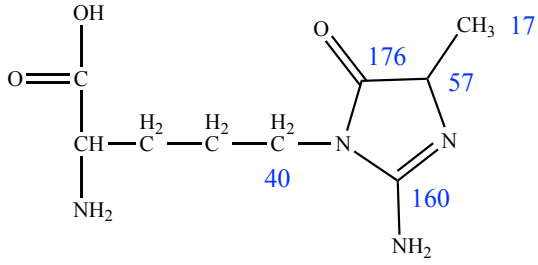
6. Methylglyoxal-hydroimidazolone (MG-H1) and ^{13}C chemical shifts [188]



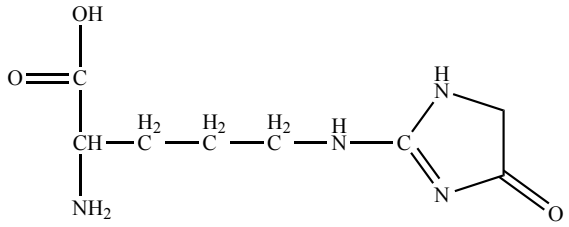
Methylglyoxal-hydroimidazolone (MG-H2) and ^{13}C chemical shifts [188]



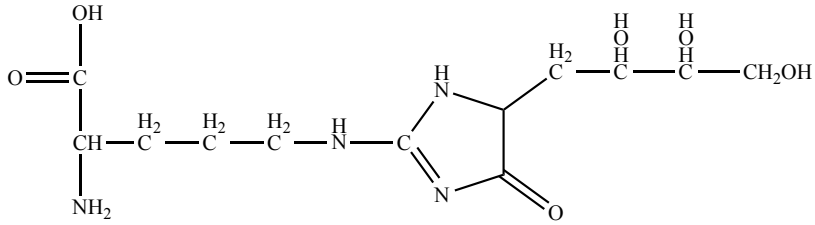
Methylglyoxal-hydroimidazolone (MG-H3) and ^{13}C chemical shifts [188]



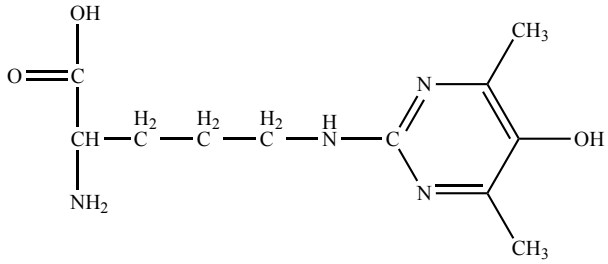
7. Glyoxal-hydroimidazolone (G-H1)



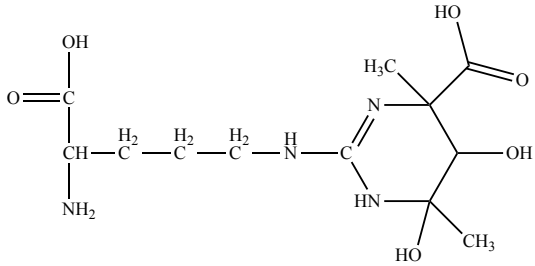
8. 3-Deoxyglucosone-hydroimidazolone (3DG-H)



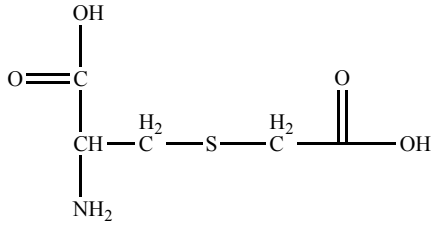
9. Argpyrimidine and ^{13}C chemical shifts



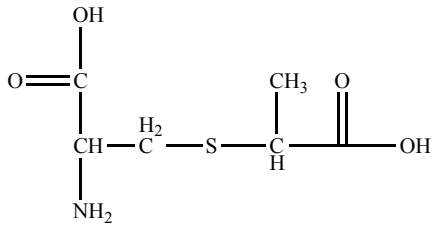
10. Tetrahydropyrimidine (THP) and ^{13}C chemical shifts



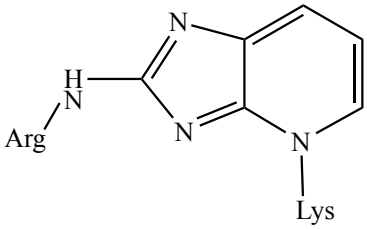
11. S-(carboxymethyl)cysteine (CMC)



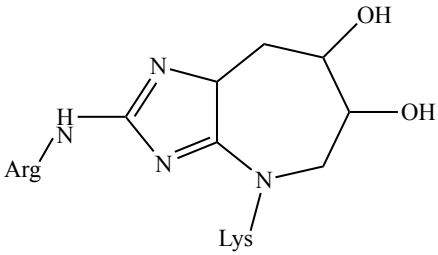
12. S-(carboxyethyl)cysteine (CEC)



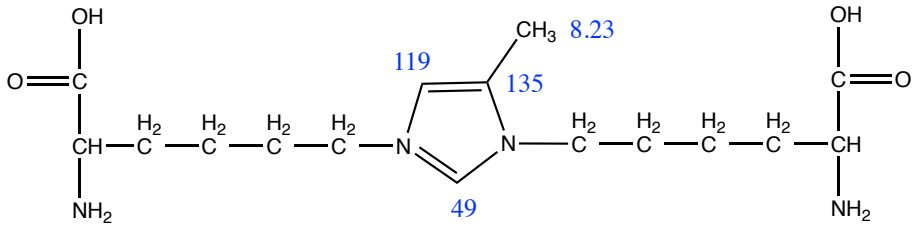
13. Pentosidine



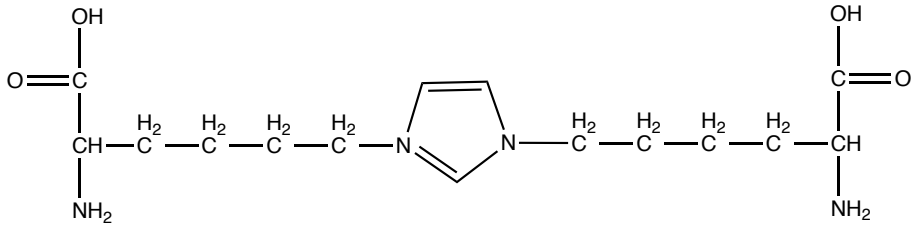
14. Glucosepane



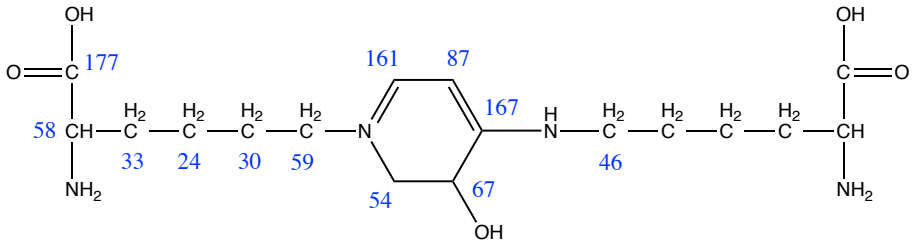
15. Methylglyoxal-lysine dimer (MOLD) (or Imidazolysine) and ^{13}C chemical shifts [189]



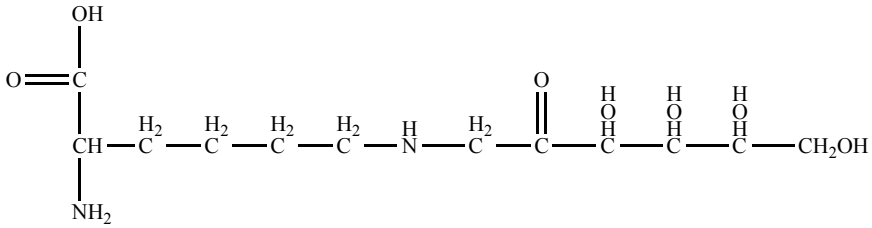
16. Glyoxal-lysine dimer (GOLD)



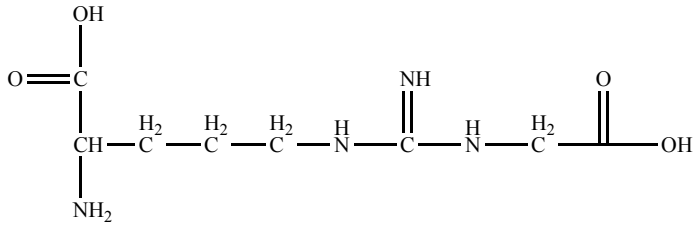
17. Lysine-dihydropyridinium (K2P) and ^{13}C chemical shifts [107]



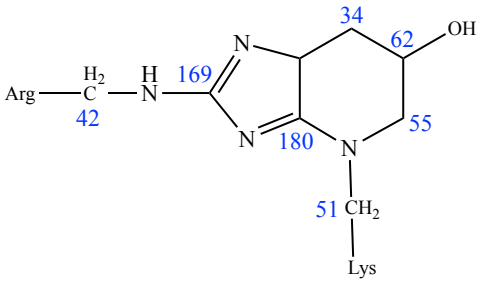
18. Fructose-lysine



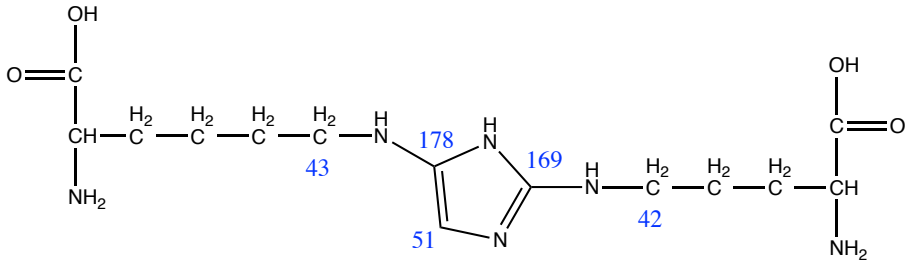
19. N-(carboxymethyl)arginine (CMA)



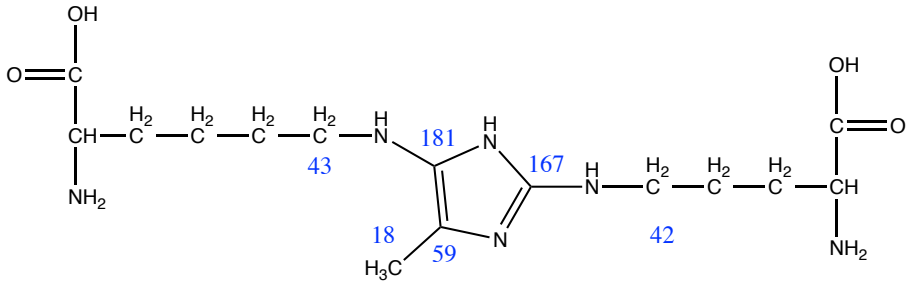
20. Pentosinane and ^{13}C chemical shifts [109]



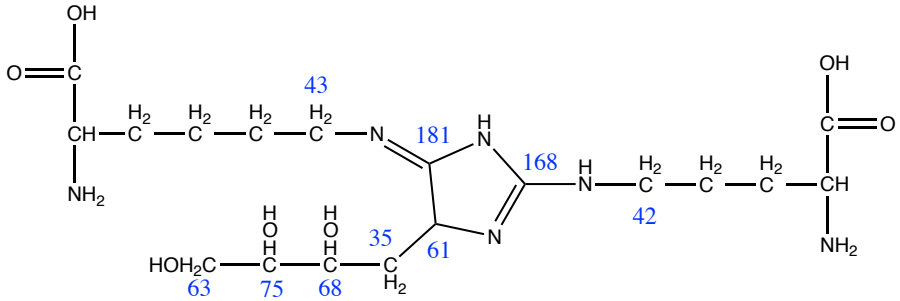
21. Glyoxal-derived-imidazoline-crosslink (GODIC) and ^{13}C chemical shifts [110]



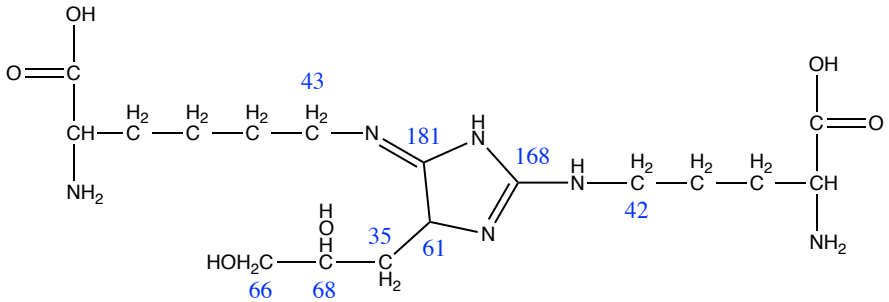
22. Methylglyoxal-derived-imidazoline-crosslink (MODIC) and ^{13}C chemical shifts [110]



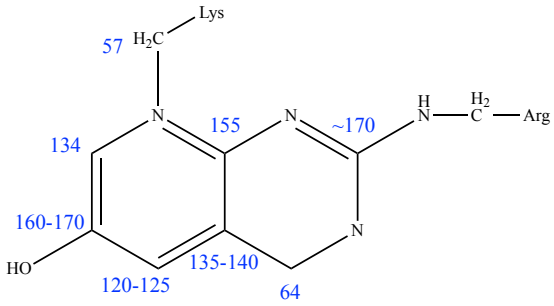
23. DOGDIC and ^{13}C chemical shifts [109]



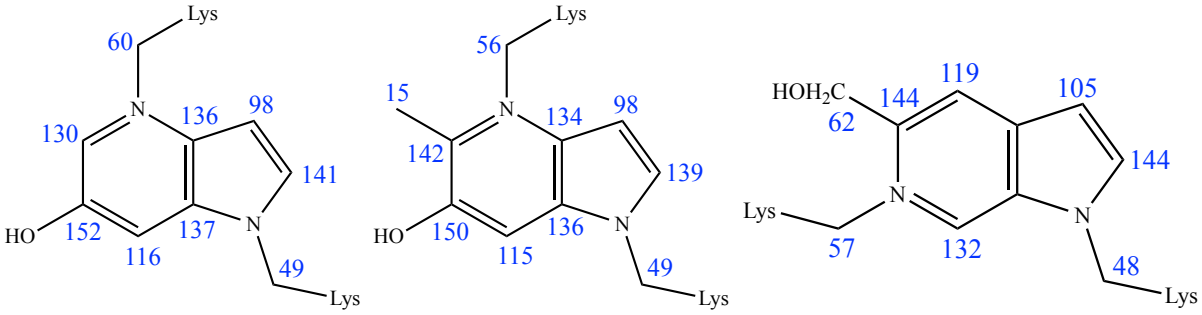
24. DOPDIC and ^{13}C chemical shifts [109]



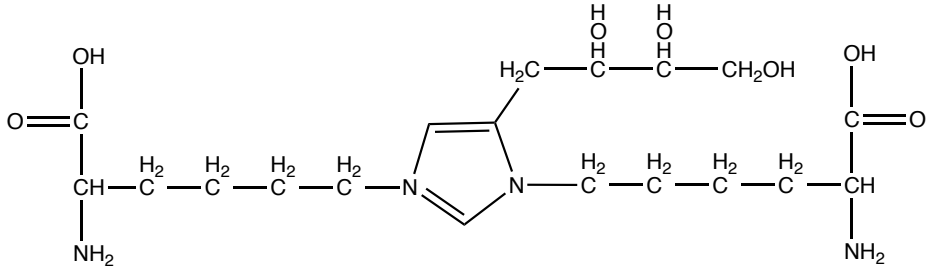
25. Arg-hydroxy-triosidine and ^{13}C chemical shifts [111]



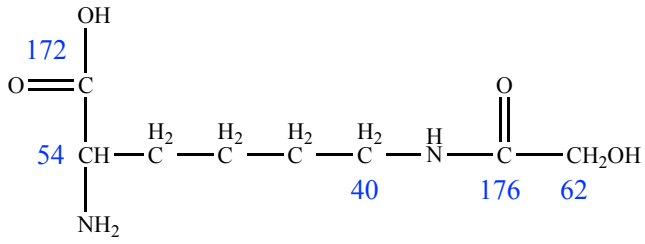
26. Vesperlysine A, B and C and ^{13}C chemical shifts [112]



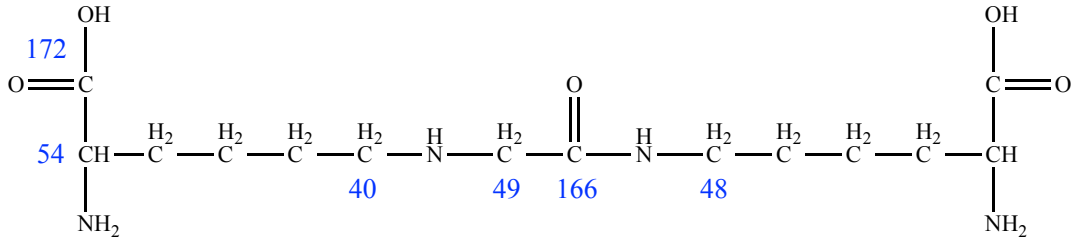
27. 3-Deoxyglucosone-lysine dimer (DOLD)



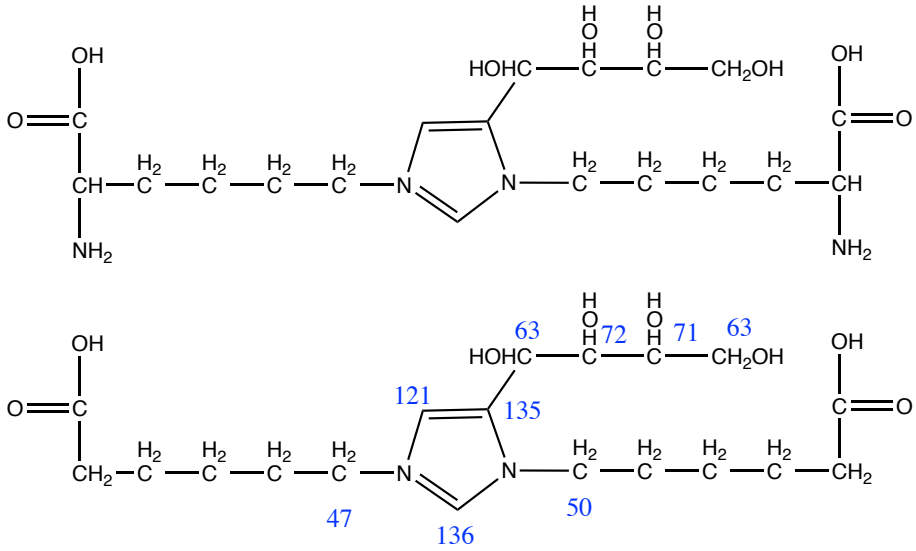
28. GALA and ^{13}C chemical shifts [114]



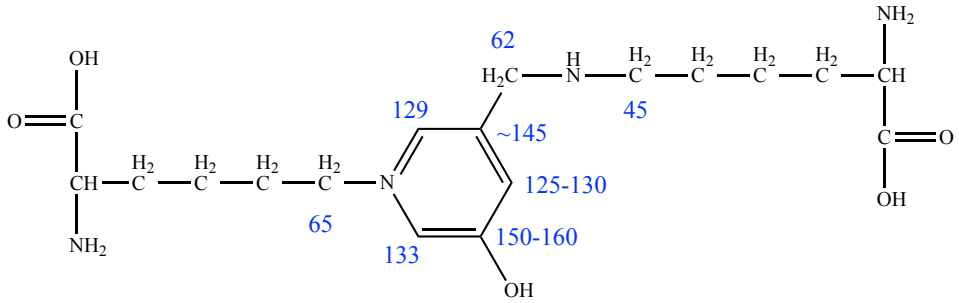
29. GOLA and ^{13}C chemical shifts [114]



30. Glucose-lysine dimer (GLUCOLD) (above) and ^{13}C chemical shifts from GLUCOLD analogue (below) [115]



31. Lys-hydroxy-triosidine and ^{13}C chemical shifts [111]



Bibliography

[1] Cell Biology of Extracellular Matrix. Springer US: Boston, MA 1991.

[2] Kini, U. and Nandeesh, B. N. Radionuclide and Hybrid Bone Imaging. **2012**, 29–57.

[3] Engel, J. and Bächinger, H. P. Collagen: Primer in Structure, Processing and Assembly. Springer Berlin Heidelberg: Berlin, Heidelberg 2005.

[4] Collagen: Structure and Mechanics. Springer US: Boston, MA 2008.

[5] Fibrous Proteins: Structures and Mechanisms. Springer International Publishing: Cham 2017.

[6] Canty, E. G. and Kadler, K. E. Collagen fibril biosynthesis in tendon: A review and recent insights. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology* **2002**, 133, 979–985.

[7] Shoulders, M. D. and Raines, R. T. Collagen Structure and Stability. *Annual Review of Biochemistry* **2009**, 78, 929–958.

[8] Ricard, S. The Collagen Family. *Cold Spring Harbor Perspectives in Biology* **2011**, 3, 1–19.

[9] Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; and Darnell, J. Molecular Cell Biology. 4th edition. W.H. Freeman and Co: New York 2000.

[10] Mouw, J. K. , Ou, G. , and Weaver, V. M. Extracellular matrix assembly: a multiscale deconstruction. *Nature Reviews Molecular Cell Biology* **2014**, 15, 771–785.

[11] Bella, J. Collagen structure: new tricks from a very old dog. *Biochemical Journal* **2016**, 473, 1001–1025.

[12] Ramshaw, J. A. . , Shah, N. K. , and Brodsky, B. Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host–Guest Triple-Helical Peptides. *Journal of Structural Biology* **1998**, 122, 86–91.

[13] Brodsky, B. , Thiagarajan, G. , Madhan, B. , and Kar, K. Triple-helical peptides: An approach to collagen conformation, stability, and self-association. *Biopolymers* **2008**, 89, 345–353.

[14] Fallas, J. A. , Dong, J. , Tao, Y. J. , and Hartgerink, J. D. Structural Insights into Charge Pair Interactions in Triple Helical Collagen-like Proteins. *Journal of Biological Chemistry* **2012**, 287, 8039–8047.

[15] Ramachandran, G. X. and Chandrasekharan, R. Interchain hydrogen bonds via bound

water molecules in the collagen triple helix. *Biopolymers* **1968**, 6, 1649–1658.
[16] Orgel, J. P. R. O. , Persikov, A. V. , and Antipova, O. Variation in the Helical Structure of Native Collagen. *PLoS ONE* **2014**, 9, e89519.
[17] Malhotra, V. and Erlmann, P. The Pathway of Collagen Secretion. *Annual Review of Cell and Developmental Biology* **2015**, 31, 109–124.
[18] Ferreira, A. M. , Gentile, P. , Chiono, V. , and Ciardelli, G. Collagen for bone tissue regeneration. *Acta Biomaterialia* **2012**, 8, 3191–3200.
[19] Orgel, J. P. , Wess, T. J. , and Miller, A. The in situ conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen. *Structure* **2000**, 8, 137–142.
[20] MILLER, A. and WRAY, J. S. Molecular Packing in Collagen. *Nature* **1971**, 230, 437–439.
[21] Hulmes, D. J. S. , Miller, A. , Parry, D. A. D. , Piez, K. A. , and Woodhead-Galloway, J. Analysis of the primary structure of collagen for the origins of molecular packing. *Journal of Molecular Biology* **1973**, 79, 137–148.
[22] Bailey, A. J. , Paul, R. G. , and Knott, L. Mechanisms of maturation and ageing of collagen. *Mechanisms of Ageing and Development* **1998**, 106, 1–56.
[23] Fleischmajer, R. , Douglas MacDonald, E. , Perlish, J. S. , Burgeson, R. E. , and Fisher, L. W. Dermal collagen fibrils are hybrids of type I and type III collagen molecules. *Journal of Structural Biology* **1990**, 105, 162–169.
[24] Xu, J. and Shi, G.-P. Vascular wall extracellular matrix proteins and vascular diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2014**, 1842, 2106–2119.
[25] Wenstrup, R. J. , Smith, S. M. , Florer, J. B. , Zhang, G. , Beason, D. P. , Seegmiller, R. E. , Soslowsky, L. J. , and Birk, D. E. Regulation of collagen fibril nucleation and initial fibril assembly involves coordinate interactions with collagens V and XI in developing tendon. *Journal of Biological Chemistry* **2011**, 286, 20455–20465.
[26] Humphries, J. D. , Byron, A. , and Humphries, M. J. Integrin ligands at a glance. *Journal of Cell Science* **2006**, 119, 3901–3903.
[27] Bonnans, C. , Chou, J. , and Werb, Z. Remodelling the extracellular matrix in development and disease. *Nature Reviews Molecular Cell Biology* **2014**, 15, 786–801.
[28] Järveläinen, H. , Sainio, A. , Koulu, M. , Wight, T. N. , and Penttinen, R. Extracellular Matrix Molecules: Potential Targets in Pharmacotherapy. *Pharmacological Reviews* **2009**, 61, 198–223.
[29] Bailey, A. Molecular mechanisms of ageing in connective tissues. *Mechanisms of Ageing*

and Development **2001**, 122, 735–755.

[30] Law, B. , Fowlkes, V. , Goldsmith, J. G. , Carver, W. , and Goldsmith, E. C. Diabetes-Induced Alterations in the Extracellular Matrix and Their Impact on Myocardial Function. *Microscopy and Microanalysis* **2012**, 18, 22–34.

[31] Butcher, D. T. , Alliston, T. , and Weaver, V. M. A tense situation: forcing tumour progression. *Nature Reviews Cancer* **2009**, 9, 108–122.

[32] Mouw, J. K. , Yui, Y. , Damiano, L. , Bainer, R. O. , Lakins, J. N. , Acerbi, I. , Ou, G. , Wijekoon, A. C. , Levental, K. R. , Gilbert, P. M. , Hwang, E. S. , Chen, Y.-Y. , and Weaver, V. M. Tissue mechanics modulate microRNA-dependent PTEN expression to regulate malignant progression. *Nature Medicine* **2014**, 20, 360–367.

[33] Levental, K. R. , Yu, H. , Kass, L. , Lakins, J. N. , Egeblad, M. , Erler, J. T. , Fong, S. F. T. , Csiszar, K. , Giaccia, A. , Weninger, W. , Yamauchi, M. , Gasser, D. L. , and Weaver, V. M. Matrix Crosslinking Forces Tumor Progression by Enhancing Integrin Signaling. *Cell* **2009**, 139, 891–906.

[34] Frantz, C. , Stewart, K. M. , and Weaver, V. M. The extracellular matrix at a glance. *Journal of Cell Science* **2010**, 123, 4195–4200.

[35] Iozzo, R. V. and Schaefer, L. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biology* **2015**, 42, 11–55.

[36] Schaefer, L. and Schaefer, R. M. Proteoglycans: from structural compounds to signaling molecules. *Cell and Tissue Research* **2010**, 339, 237–246.

[37] Couchman, J. R. and Pataki, C. A. An Introduction to Proteoglycans and Their Localization. *Journal of Histochemistry & Cytochemistry* **2012**, 60, 885–897.

[38] Pomin, V. and Mulloy, B. Glycosaminoglycans and Proteoglycans. *Pharmaceuticals* **2018**, 11, 27.

[39] Häcker, U. , Nybakken, K. , and Perrimon, N. Heparan sulphate proteoglycans: The sweet side of development. *Nature Reviews Molecular Cell Biology* **2005**, 6, 530–541.

[40] FISCHER, G. M. and LLAURADO, J. G. Collagen and Elastin Content in Canine Arteries Selected from Functionally Different Vascular Beds. *Circulation Research* **1966**, 19, 394–399.

[41] Green, E. M. , Mansfield, J. C. , Bell, J. S. , and Winlove, C. P. The structure and micromechanics of elastic tissue. *Interface Focus* **2014**, 4

[42] Kielty, C. M. , Sherratt, M. J. , and Shuttleworth, C. A. Elastic fibres. *Journal of Cell Science* **2002**, 115, 2817 LP – 2828.

[43] Schwarzbauer, J. E. and DeSimone, D. W. Fibronectins, their fibrillogenesis, and in vivo functions. *Cold Spring Harbor Perspectives in Biology* **2011**, 3, 1–19.

[44] Hynes, R. O. and Naba, A. Overview of the Matrisome--An Inventory of Extracellular Matrix Constituents and Functions. *Cold Spring Harbor Perspectives in Biology* **2012**, 4, a004903–a004903.

[45] Monnier, V. and Cerami, A. Nonenzymatic browning in vivo: possible process for aging of long-lived proteins. *Science* **1981**, 211, 491–493.

[46] Cerami, A. Hypothesis: Glucose as a Mediator of Aging. *Journal of the American Geriatrics Society* **1985**, 33, 626–634.

[47] Kennedy, L. and Baynes, J. W. Non-enzymatic glycosylation and the chronic complications of diabetes: an overview. *Diabetologia* **1984**, 26, 93–98.

[48] Monnier, V. M. , Kohn, R. R. , and Cerami, A. Accelerated age-related browning of human collagen in diabetes mellitus. *Proceedings of the National Academy of Sciences* **1984**, 81, 583–587.

[49] Sell, D. R. and Monnier, V. M. End-stage renal disease and diabetes catalyze the formation of a pentose-derived crosslink from aging human collagen. *Journal of Clinical Investigation* **1990**, 85, 380–384.

[50] Singh, V. P. , Bali, A. , Singh, N. , and Jaggi, A. S. Advanced Glycation End Products and Diabetic Complications. *The Korean Journal of Physiology & Pharmacology* **2014**, 18, 1.

[51] Negre-Salvayre, A. , Salvayre, R. , Augé, N. , Pamplona, R. , and Portero-Otín, M. Hyperglycemia and Glycation in Diabetic Complications. *Antioxidants & Redox Signaling* **2009**, 11, 3071–3109.

[52] Ahmed, N. Advanced glycation endproducts—role in pathology of diabetic complications. *Diabetes Research and Clinical Practice* **2005**, 67, 3–21.

[53] Smith, M. A. , Taneda, S. , Richey, P. L. , Miyata, S. , Yan, S. D. , Stern, D. , Sayre, L. M. , Monnier, V. M. , and Perry, G. Advanced Maillard reaction end products are associated with Alzheimer disease pathology. *Proceedings of the National Academy of Sciences* **1994**, 91, 5710–5714.

[54] Sasaki, N. , Fukatsu, R. , Tsuzuki, K. , Hayashi, Y. , Yoshida, T. , Fujii, N. , Koike, T. , Wakayama, I. , Yanagihara, R. , Garruto, R. , Amano, N. , and Makita, Z. Advanced Glycation End Products in Alzheimer's Disease and Other Neurodegenerative Diseases. *The American Journal of Pathology* **1998**, 153, 1149–1155.

[55] Yan, S. Du , Chen, X. , Fu, J. , Chen, M. , Zhu, H. , Roher, A. , Slattery, T. , Zhao, L. , Nagashima, M. , Morser, J. , Migheli, A. , Nawroth, P. , Stern, D. , and Schmidt, A. M. RAGE and amyloid-β peptide neurotoxicity in Alzheimer's disease. *Nature* **1996**, 382, 685–691.

[56] Choi, B.-R. , Cho, W.-H. , Kim, J. , Lee, H. J. , Chung, C. , Jeon, W. K. , and Han, J.-S.

Increased expression of the receptor for advanced glycation end products in neurons and astrocytes in a triple transgenic mouse model of Alzheimer's disease. *Experimental & Molecular Medicine* **2014**, 46, e75–e75.

[57] Ansari, N. A. and Rasheed, Z. Non-enzymatic glycation of proteins: From diabetes to cancer. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry* **2009**, 3, 335–342.

[58] Turner, D. P. Advanced Glycation End-Products: A Biological Consequence of Lifestyle Contributing to Cancer Disparity. *Cancer Research* **2015**, 75, 1925–1929.

[59] Foster, D. , Spruill, L. , Walter, K. R. , Nogueira, L. M. , Fedarovich, H. , Turner, R. Y. , Ahmed, M. , Salley, J. D. , Ford, M. E. , Findlay, V. J. , and Turner, D. P. AGE Metabolites: A Biomarker Linked to Cancer Disparity? *Cancer Epidemiology Biomarkers & Prevention* **2014**, 23, 2186–2191.

[60] Maillard, L. C. Action of amino acids on sugars. Formation of melanoidins in a methodical way. *Compte-Rendu de l'Academie des Science* **1912**, 154, 66–68.

[61] Hodge, J. E. Dehydrated Foods, Chemistry of Browning Reactions in Model Systems. *Journal of Agricultural and Food Chemistry* **1953**, 1, 928–943.

[62] Mitsuda, H. , Yasumoto, K. , and Yokoyama, K. Studies on the Free Radical in Amino-Carbonyl Reaction. *Agricultural and Biological Chemistry* **1965**, 29, 751–756.

[63] Namiki, M. , Hayashi, T. , and Kawakishi, S. Free Radicals Developed in the Amino-Garbonyl Reaction of Sugars with Amino Acids. *Agricultural and Biological Chemistry* **1973**, 37, 2935–2936.

[64] Namiki, M. and Hayashi, T. Formation of novel free radical products in an early stage of Maillard reaction. *Progress in food and nutrition science* **1981**, 5, 81–91.

[65] Slatter, D. A. , Avery, N. C. , and Bailey, A. J. Identification of a New Cross-link and Unique Histidine Adduct from Bovine Serum Albumin Incubated with Malondialdehyde. *Journal of Biological Chemistry* **2004**, 279, 61–69.

[66] Kita, K. , Kawashima, Y. , Makino, R. , Namauo, T. , Ogawa, S. , Muraoka, H. , and Fujimura, S. Detection of Two Types of Glycated Tryptophan Compounds in the Plasma of Chickens Fed Tryptophan Excess Diets. *The Journal of Poultry Science* **2013**, 50, 138–142.

[67] Thorpe, S. R. and Baynes, J. W. Maillard reaction products in tissue proteins: New products and new perspectives. *Amino Acids* **2003**, 25, 275–281.

[68] Mostafa, A. A. , Randell, E. W. , Vasdev, S. C. , Gill, V. D. , Han, Y. , Gadag, V. , Raouf, A. A. , and Said, H. El Plasma protein advanced glycation end products, carboxymethyl cysteine, and carboxyethyl cysteine, are elevated and related to nephropathy in patients with diabetes. *Molecular and Cellular Biochemistry* **2007**, 302, 35–42.

[69] Nursten, H. E. The Maillard Reaction: Chemistry, Biochemistry and Implications. Royal Society of Chemistry: Cambridge 2005.

[70] Odani, H. , Shinzato, T. , Matsumoto, Y. , Usami, J. , and Maeda, K. Increase in three alpha,beta-Dicarbonyl Compound Levels in Human Uremic Plasma: Specific in vivo determination of intermediates in advanced maillard reaction. *Biochemical and Biophysical Research Communications* **1999**, 93, 89–93.

[71] Paul, R. G. and Bailey, A. J. The effect of advanced glycation end-product formation upon cell-matrix interactions. *International Journal of Biochemistry and Cell Biology* **1999**, 31, 653–660.

[72] Reigle, K. L. , Lullo, G. Di , Turner, K. R. , Last, J. A. , Chervoneva, I. , Birk, D. E. , Funderburgh, J. L. , Elrod, E. , Germann, M. W. , Surber, C. , Sanderson, R. D. , and San Antonio, J. D. Non-enzymatic glycation of type I collagen diminishes collagen-proteoglycan binding and weakens cell adhesion. *Journal of Cellular Biochemistry* **2008**, 104, 1684–1698.

[73] Hudson, D. M. , Archer, M. , King, K. B. , and Eyre, D. R. Glycation of type I collagen selectively targets the same helical domain lysine sites as lysyl oxidase–mediated cross-linking. *Journal of Biological Chemistry* **2018**, 293, 15620–15627.

[74] Avery, N. C. and Bailey, A. J. The effects of the Maillard reaction on the physical properties and cell interactions of collagen. *Pathologie Biologie* **2006**, 54, 387–395.

[75] Winlove, C. P. , Parker, K. H. , Avery, N. C. , and Bailey, A. J. Interactions of elastin and aorta with sugars in vitro and their effects on biochemical and physical properties. *Diabetologia* **1996**, 39, 1131–1139.

[76] Li, Y. , Fessel, G. , Georgiadis, M. , and Snedeker, J. G. Advanced glycation end-products diminish tendon collagen fiber sliding. *Matrix Biology* **2013**, 32, 169–177.

[77] Dunn, J. A. , Patrick, J. S. , Thorpe, S. R. , and Baynes, J. W. Oxidation of glycated proteins: age-dependent accumulation of N.epsilon.-(carboxymethyl)lysine in lens proteins. *Biochemistry* **1989**, 28, 9464–9468.

[78] Dunn, J. A. , McCance, D. R. , Thorpe, S. R. , Lyons, T. J. , and Baynes, J. W. Age-dependent accumulation of N.epsilon.-(carboxymethyl)lysine and N.epsilon.-(carboxymethyl)hydroxylysine in human skin collagen. *Biochemistry* **1991**, 30, 1205–1210.

[79] Ahmedsq, M. U. , Thorpes, S. R. , and Baynesst, J. W. Identification of W-Carboxymethyllysine as a Degradation Product of Fructoselysine in Glycated Protein. **1986**, 261, 4889–4894.

[80] Sell, D. R. and Monnier, V. M. Structure elucidation of a senescence cross-link from human extracellular matrix. Implication of pentoses in the aging process. *The Journal of*

biological chemistry **1989**, 264, 21597–21602.

[81] Sell, D. R. , Biemel, K. M. , Reihl, O. , Lederer, M. O. , Strauch, C. M. , and Monnier, V. M. Glucosepane Is a Major Protein Cross-link of the Senescent Human Extracellular Matrix. *Journal of Biological Chemistry* **2005**, 280, 12310–12315.

[82] Monnier, V. M. , Kohn, R. R. , and Cerami, A. Accelerated age-related browning of human collagen in diabetes mellitus. *Proceedings of the National Academy of Sciences* **1984**, 81, 583–587.

[83] Hamlin, C. R. , Kohn, R. R. , and Luschin, J. H. Apparent Accelerated Aging of Human Collagen in Diabetes Mellitus. *Diabetes* **1975**, 24, 902–904.

[84] Fan, X. , Sell, D. R. , Zhang, J. , Nemet, I. , Theves, M. , Lu, J. , Strauch, C. , Halushka, M. K. , and Monnier, V. M. Anaerobic vs aerobic pathways of carbonyl and oxidant stress in human lens and skin during aging and in diabetes: A comparative analysis. *Free Radical Biology and Medicine* **2010**, 49, 847–856.

[85] Sugiyama, S. , Miyata, T. , Horie, K. , Iida, Y. , Tsuyuki, M. , Tanaka, H. , and Maeda, K. Advanced glycation end-products in diabetic nephropathy. *Nephrology Dialysis Transplantation* **1996**, 11, 91–94.

[86] Tanji, N. , Markowitz, G. S. , Fu, C. , Kislinger, T. , Taguchi, A. , Pischetsrieder, M. , Stern, D. , Schmidt, A. M. , and D'Agati, V. D. Expression of advanced glycation end products and their cellular receptor RAGE diabetic nephropathy and nondiabetic renal disease. *Journal of the American Society of Nephrology* **2000**, 11, 1656–1666.

[87] Zhu, Y. , Shu, T. , Lin, Y. , Wang, H. , Yang, J. , Shi, Y. , and Han, X. Inhibition of the receptor for advanced glycation endproducts (RAGE) protects pancreatic β-cells. *Biochemical and Biophysical Research Communications* **2011**, 404, 159–165.

[88] Schalkwijk, C. G. , Stehouwer, C. D. A. , and Hinsbergh, V. W. M. van Fructose-mediated non-enzymatic glycation: Sweet coupling or bad modification. *Diabetes/Metabolism Research and Reviews* **2004**, 20, 369–382.

[89] Hotta, N. , Kakuta, H. , Fukasawa, H. , Kimura, M. , Koh, N. , Iida, M. , Terashima, H. , Morimura, T. , and Sakamoto, N. Effects of a fructose-rich diet and the aldose reductase inhibitor, ONO-2235, on the development of diabetic neuropathy in streptozotocin-treated rats. *Diabetologia* **1985**, 28, 176–180.

[90] Schröter, D. and Höhn, A. Role of Advanced Glycation End Products in Carcinogenesis and their Therapeutic Implications. *Current Pharmaceutical Design* **2019**, 24, 5245–5251.

[91] Cairns, R. A. , Harris, I. S. , and Mak, T. W. Regulation of cancer cell metabolism. *Nature Reviews Cancer* **2011**, 11, 85–95.

[92] Riehl, A. , Németh, J. , Angel, P. , and Hess, J. The receptor RAGE: Bridging inflammation and cancer. *Cell Communication and Signaling* **2009**, 7, 1–7.
[93] Atsumi, T. Diabetes and risk of cancer. *Diabetology International* **2015**, 6, 190–192.
[94] Rojas, A. , Añazco, C. , González, I. , and Araya, P. Extracellular matrix glycation and receptor for advanced glycation end-products activation: a missing piece in the puzzle of the association between diabetes and cancer. *Carcinogenesis* **2018**, 39, 515–521.
[95] Shlomai, G. , Neel, B. , LeRoith, D. , and Gallagher, E. J. Type 2 Diabetes Mellitus and Cancer: The Role of Pharmacotherapy. *Journal of Clinical Oncology* **2016**, 34, 4261–4269.
[96] Rahbar, S. , Blumenfeld, O. , and Ranney, H. M. Studies of an unusual hemoglobin in patients with diabetes mellitus. *Biochemical and Biophysical Research Communications* **1969**, 36, 838–843.
[97] Monnier, V. M. , Sun, W. , Sell, D. R. , Fan, X. , Nemet, I. , and Genuth, S. Glucosepane: a poorly understood advanced glycation end product of growing importance for diabetes and its complications. *Clinical Chemistry and Laboratory Medicine* **2014**, 52
[98] THORNALLEY, P. J. , BATTAH, S. , AHMED, N. , KARACHALIAS, N. , AGALOU, S. , BABAEI-JADIDI, R. , and DAWNAY, A. Quantitative screening of advanced glycation endproducts in cellular and extracellular proteins by tandem mass spectrometry. *Biochemical Journal* **2003**, 375, 581–592.
[99] Miyata, S. and Monnier, V. Immunohistochemical detection of advanced glycosylation end products in diabetic tissues using monoclonal antibody to pyrraline. *Journal of Clinical Investigation* **1992**, 89, 1102–1112.
[100] Nagai, R. , Hayashi, C. M. , Xia, L. , Takeya, M. , and Horiuchi, S. Identification in human atherosclerotic lesions of GA-pyridine, a novel structure derived from glycolaldehyde-modified proteins. *Journal of Biological Chemistry* **2002**, 277, 48905–48912.
[101] Oya, T. , Hattori, N. , Mizuno, Y. , Miyata, S. , Maeda, S. , Osawa, T. , and Uchida, K. Methylglyoxal Modification of Protein. *Journal of Biological Chemistry* **1999**, 274, 18492–18502.
[102] Eupen, M. G. A. van , Schram, M. T. , Colhoun, H. M. , Hanssen, N. M. J. , Niessen, H. W. M. , Tarnow, L. , Parving, H. H. , Rossing, P. , Stehouwer, C. D. A. , and Schalkwijk, C. G. The methylglyoxal-derived AGE tetrahydropyrimidine is increased in plasma of individuals with type 1 diabetes mellitus and in atherosclerotic lesions and is associated with sVCAM-1. *Diabetologia* **2013**, 56, 1845–1855.
[103] Mostafa, A. A. , Randell, E. W. , Vasdev, S. C. , Gill, V. D. , Han, Y. , Gadag, V. , Raouf, A. A. , and Said, H. El Plasma protein advanced glycation end products, carboxymethyl

cysteine, and carboxyethyl cysteine, are elevated and related to nephropathy in patients with diabetes. *Molecular and Cellular Biochemistry* **2007**, 302, 35–42.

[104] Naudí, A. , Jové, M. , Cacabelos, D. , Ayala, V. , Cabre, R. , Caro, P. , Gomez, J. , Portero-Otín, M. , Barja, G. , and Pamplona, R. Formation of S-(carboxymethyl)-cysteine in rat liver mitochondrial proteins: effects of caloric and methionine restriction. *Amino Acids* **2013**, 44, 361–371.

[105] D.R., S. , R.H., N. , S.K., G. , P., O. , A., L. , and J., F. Pentosidine: A molecular marker for cumulative damage to proteins in diabetes, aging, and uremia. *Diabetes/Metabolism Reviews* **1991**, 7, 239–251.

[106] Frye, E. B. , Degenhardt, T. P. , Thorpe, S. R. , and Baynes, J. W. Role of the Maillard Reaction in Aging of Tissue Proteins. *Journal of Biological Chemistry* **1998**, 273, 18714–18719.

[107] Cheng, R. , Feng, Q. , Argirov, O. K. , and Ortwerth, B. J. Structure elucidation of a novel yellow chromophore from human lens protein. *Journal of Biological Chemistry* **2004**, 279, 45441–45449.

[108] Iijima, K. , Murata, M. , Takahara, H. , Irie, S. , and Fujimoto, D. Identification of N-carboxymethylarginine as a novel acid labile advanced glycation end product in collagen. *Biochem. J.* **2000**, 347, 23–27.

[109] Biemel, K. M. , Reihl, O. , Conrad, J. , and Lederer, M. O. Formation pathways for lysine-arginine cross-links derived from hexoses and pentoses by maillard processes: Unraveling the structure of a pentosidine precursor. *Journal of Biological Chemistry* **2001**, 276, 23405–23412.

[110] Lederer, M. O. and Bühler, H. P. Cross-linking of proteins by maillard processes - Characterization and detection of a lysine-arginine cross-link derived from D-glucose. *Bioorganic and Medicinal Chemistry* **1999**, 7, 1081–1088.

[111] TESSIER, F. J. , MONNIER, V. M. , SAYRE, L. M. , and KORNFIELD, J. A. Triosidines: novel Maillard reaction products and cross-links from the reaction of triose sugars with lysine and arginine residues. *Biochemical Journal* **2003**, 369, 705–719.

[112] Nakamura, K. , Nakazawa, Y. , and Ienaga, K. Acid-stable fluorescent advanced glycation end products: Vesperlysines A, B, and C are formed as crosslinked products in the maillard reaction between lysine or proteins with glucose. *Biochemical and Biophysical Research Communications* **1997**, 232, 227–230.

[113] Skovsted, I. C.; Christensen, M.; and Mortensen, S. B. Characterisation of a Novel AGE-Epitope Derived from Lysine and 3-Deoxyglucosone. *In The Maillard Reaction in Foods and*

Medicine; Elsevier, 2005; p. 444.
[114] Glomb, M. A. and Pfahler, C. Amides are novel protein modifications formed by physiological sugars. *International Congress Series* **2002**, 1245, 501–502.
[115] Nemet, I. , Strauch, C. M. , and Monnier, V. M. Favored and disfavored pathways of protein crosslinking by glucose: glucose lysine dimer (GLUCOLD) and crossline versus glucosepane. *Amino Acids* **2011**, 40, 167–181.
[116] Thornalley, P. J. Measurement of Protein Glycation, Glycated Peptides and Glycation Free Adducts. *Peritoneal Dialysis International* **2005**, 25, 522–533.
[117] Lapolla, A. , Traldi, P. , and Fedele, D. Importance of measuring products of non-enzymatic glycation of proteins. *Clinical Biochemistry* **2005**, 38, 103–115.
[118] Thornalley, P. J. and Rabbani, N. Detection of oxidized and glycated proteins in clinical samples using mass spectrometry — A user's perspective. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2014**, 1840, 818–829.
[119] Ashraf, J. M. , Ahmad, S. , Choi, I. , Ahmad, N. , Farhan, M. , Tatyana, G. , and Shahab, U. Recent advances in detection of AGEs: Immunochemical, bioanalytical and biochemical approaches. *IUBMB Life* **2015**, 67, 897–913.
[120] Chow, W. Y. , Rajan, R. , Muller, K. H. , Reid, D. G. , Skepper, J. N. , Wong, W. C. , Brooks, R. A. , Green, M. , Bihan, D. , Farndale, R. W. , Slatter, D. A. , Shanahan, C. M. , and Duer, M. J. NMR Spectroscopy of Native and in Vitro Tissues Implicates PolyADP Ribose in Biomineralization. *Science* **2014**, 344, 742–746.
[121] Bunn, H. F. and Higgins, P. J. Reaction of monosaccharides with proteins: Possible evolutionary significance. *Science* **1981**, 213, 222–224.
[122] Keeler, J. H. Understanding NMR spectroscopy. John Wiley & Sons.: Chichester 2005.
[123] Duer, M. J. Introduction to Solid-State NMR Spectroscopy. Wiley-Blackwell: Oxford 2004.
[124] Apperley, D. C.; Harris, R. K.; and Hodgkinson, P. Solid-State NMR : basic principles & practice. Momentum: New York 2012.
[125] Bennett, A. E. , Rienstra, C. M. , Auger, M. , Lakshmi, K. V. , and Griffin, R. G. Heteronuclear decoupling in rotating solids. *The Journal of Chemical Physics* **1995**, 103, 6951–6958.
[126] Fung, B. M. , Khitrin, A. K. , and Ermolaev, K. An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids. *Journal of Magnetic Resonance* **2000**, 142, 97–101.
[127] Hartmann, S. R. and Hahn, E. L. Nuclear Double Resonance in the Rotating Frame. *Physical Review* **1962**, 128, 2042–2053.

[128] G, M., X, W., and O, S.S. Ramped-Amplitude Cross Polarization in Magic-Angle-Spinning NMR. *Journal of Magnetic Resonance Series A*. , 110, 219–227. (1994).
[129] Carver, T. R. and Slichter, C. P. Polarization of Nuclear Spins in Metals. *Physical Review* **1953**, 92, 212–213.
[130] Overhauser, A. W. Polarization of Nuclei in Metals. *Physical Review* **1953**, 92, 411–415.
[131] Sauvée, C. , Rosay, M. , Casano, G. , Aussenac, F. , Weber, R. T. , Ouari, O. , and Tordo, P. Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angewandte Chemie International Edition* **2013**, 52, 10858–10861.
[132] Szeverenyi, N. M. , Sullivan, M. J. , and Maciel, G. E. Observation of spin exchange by two-dimensional fourier transform 13C cross polarization-magic-angle spinning. *Journal of Magnetic Resonance (1969)* **1982**, 47, 462–475.
[133] Takegoshi, K. , Nakamura, S. , and Terao, T. 13C–1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chemical Physics Letters* **2001**, 344, 631–637.
[134] Foerster, H., Struppe, J., Steuernagel, S., Aussenacc, F., Benevelli, F., Gierth, P., and Wegner, S. Solid State NMR AVANCE Solids User Manual. (2009).
[135] Chow, W. Y. A Solid-State NMR Toolkit for Probing Collagen Atomic Structure in the Extracellular Matrix, 2013.
[136] Wong, W. C. Elucidating Glycation-Induced Structural Changes in Collagen by Solid-State NMR, 2016.
[137] Lee, Y. . , Kurur, N. . , Helmle, M. , Johannessen, O. . , Nielsen, N. . , and Levitt, M. . Efficient dipolar recoupling in the NMR of rotating solids. A sevenfold symmetric radiofrequency pulse sequence. *Chemical Physics Letters* **1995**, 242, 304–309.
[138] Hohwy, M. , Jakobsen, H. J. , Edén, M. , Levitt, M. H. , and Nielsen, N. C. Broadband dipolar recoupling in the nuclear magnetic resonance of rotating solids: A compensated C7 pulse sequence. *Journal of Chemical Physics* **1998**, 108, 2686–2694.
[139] Hohwy, M. , Rienstra, C. M. , Jaroniec, C. P. , and Griffin, R. G. Fivefold symmetric homonuclear dipolar recoupling in rotating solids: Application to double quantum spectroscopy. *Journal of Chemical Physics* **1999**, 110, 7983–7992.
[140] Hong, M. Solid-State Dipolar INADEQUATE NMR Spectroscopy with a Large Double-Quantum Spectral Width. *Journal of Magnetic Resonance* **1999**, 136, 86–91.
[141] Rossum, B.-J. van , Förster, H. , and Groot, H. J. M. de High-Field and High-Speed CP-MAS13C NMR Heteronuclear Dipolar-Correlation Spectroscopy of Solids with Frequency-Switched Lee–Goldburg Homonuclear Decoupling. *Journal of Magnetic Resonance* **1997**, 124, 516–519.

[142] Bielecki, A. , Kolbert, A. C. , and Levitt, M. H. Frequency-switched pulse sequences: Homonuclear decoupling and dilute spin NMR in solids. *Chemical Physics Letters* **1989**, 155, 341–346.

[143] Schaefer, J. , McKay, R. A. , and Stejskal, E. O. Double-cross-polarization NMR of solids. *Journal of Magnetic Resonance (1969)* **1979**, 34, 443–447.

[144] Baldus, M. , Petkova, A. T. , Herzfeld, J. , and Griffin, R. G. Cross polarization in the tilted frame: Assignment and spectral simplification in heteronuclear spin systems. *Molecular Physics* **1998**, 95, 1197–1207.

[145] Knight, C. G. , Morton, L. F. , Onley, D. J. , Peachey, A. R. , Messent, A. J. , Smethurst, P. A. , Tuckwell, D. S. , Farndale, R. W. , and Barnes, M. J. Identification in Collagen Type I of an Integrin α 2 β 1 -binding Site Containing an Essential GER Sequence. *Journal of Biological Chemistry* **1998**, 273, 33287–33294.

[146] Lee, M. and Goldburg, W. I. Nuclear-magnetic-resonance line narrowing by a rotating rf field. *Physical Review* **1965**, 140

[147] Han, J. , Tschernutter, V. , Yang, J. , Eckle, T. , and Borchers, C. H. Analysis of Selected Sugars and Sugar Phosphates in Mouse Heart Tissue by Reductive Amination and Liquid Chromatography-Electrospray Ionization Mass Spectrometry. *Analytical Chemistry* **2013**, 85, 5965–5973.

[148] Burrell, M. M. Enzymes of Molecular Biology. Humana Press: New Jersey 1993.

[149] Mótyán, J. , Tóth, F. , and Tőzsér, J. Research Applications of Proteolytic Enzymes in Molecular Biology. *Biomolecules* **2013**, 3, 923–942.

[150] Hess, G. P. , McConn, J. , Ku, E. , and McConkey, G. Studies of the activity of chymotrypsin. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **1970**, 257, 89–104.

[151] Appel, W. Chymotrypsin: Molecular and catalytic properties. *Clinical Biochemistry* **1986**, 19, 317–322.

[152] Burrell, M. M. Enzymes of Molecular Biology. Humana Press: New Jersey 1993.

[153] UniProt. https://www.uniprot.org

[154] Gilbert, T. W. , Sellaro, T. L. , and Badylak, S. F. Decellularization of tissues and organs. *Biomaterials* **2006**, 27, 3675–3683.

[155] Crapo, P. M. , Gilbert, T. W. , and Badylak, S. F. An overview of tissue and whole organ decellularization processes. *Biomaterials* **2011**, 32, 3233–3243.

[156] Principles of Fluorescence Spectroscopy. Springer US: Boston, MA 2006.

[157] Weert, M. van de and Stella, L. Fluorescence quenching and ligand binding: A critical

discussion of a popular methodology. *Journal of Molecular Structure* **2011**, 998, 144–150.
[158] Royer, C. A. Probing Protein Folding and Conformational Transitions with Fluorescence. *Chemical Reviews* **2006**, 106, 1769–1784.
[159] Pundak, S. and Roche, R. S. Tyrosine and tyrosinate fluorescence of bovine testes calmodulin: calcium and pH dependence. *Biochemistry* **1984**, 23, 1549–1555.
[160] Vivian, J. T. and Callis, P. R. Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophysical Journal* **2001**, 80, 2093–2109.
[161] Johnson, M. Detergents: Triton X-100, Tween-20, and More. *Materials and Methods* **2013**, 3
[162] Fritzsching, K. J. , Yang, Y. , Schmidt-Rohr, K. , and Hong, M. Practical use of chemical shift databases for protein solid-state NMR: 2D chemical shift maps and amino-acid assignment with secondary-structure information. *Journal of Biomolecular NMR* **2013**, 56, 155–167.
[163] Bullock, P. T. B. , Reid, D. G. , Chow, W. Y. , Lau, W. P. W. , and Duer, M. J. A new glycation product "norpronyl-lysine" and direct characterization of cross linking and other glycation adducts: NMR of model compounds and collagen. *Bioscience reports* **2014**
[164] Management of Diabetes and Hyperglycemia in Hospitalized Patients. https://www.ncbi.nlm.nih.gov/books/NBK279093/
[165] Swamy, M. S. , Tsai, C. , Abraham, A. , and Abraham, E. C. Glycation Mediated Lens Crystallin Aggregation and Cross-linking by Various Sugars and Sugar Phosphates In Vitro. *Experimental Eye Research* **1993**, 56, 177–185.
[166] Bubb, W. A. NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. *Concepts in Magnetic Resonance* **2003**, 19A, 1–19.
[167] Bailey, A. J. , Sims, T. J. , Avery, N. C. , and Halligan, E. P. Non-enzymic glycation of fibrous collagen: reaction products of glucose and ribose. *Biochemical Journal* **1995**, 305, 385–390.
[168] Clark, P. M. , Flores, G. , Evdokimov, N. M. , McCracken, M. N. , Chai, T. , Nair-Gill, E. , O'Mahony, F. , Beaven, S. W. , Faull, K. F. , Phelps, M. E. , Jung, M. E. , and Witte, O. N. Positron emission tomography probe demonstrates a striking concentration of ribose salvage in the liver. *Proceedings of the National Academy of Sciences* **2014**, 111, E2866–E2874.
[169] Munanairi, A. , O'Banion, S. K. , Gamble, R. , Breuer, E. , Harris, A. W. , and Sandwick, R. K. The multiple Maillard reactions of ribose and deoxyribose sugars and sugar phosphates. *Carbohydrate Research* **2007**, 342, 2575–2592.
[170] Drew, K. N. , Zajicek, J. , Bondo, G. , Bose, B. , and Serianni, A. S. 13C-labeled

aldopentoses: detection and quantitation of cyclic and acyclic forms by heteronuclear 1D and 2D NMR spectroscopy. *Carbohydrate Research* **1998**, 307, 199–209.

[171] Slatter, D. A. , Avery, N. C. , and Bailey, A. J. Collagen in its fibrillar state is protected from glycation. *The International Journal of Biochemistry & Cell Biology* **2008**, 40, 2253–2263.

[172] Chow, W. Y. , Li, R. , Goldberga, I. , Reid, D. G. , Rajan, R. , Clark, J. , Oschkinat, H. , Duer, M. J. , Hayward, R. , and Shanahan, C. M. Essential but sparse collagen hydroxylysyl post-translational modifications detected by DNP NMR. *Chemical Communications* **2018**, 54, 12570–12573.

[173] Delatour, T. , Fenaille, F. , Parisod, V. , Arce Vera, F. , and Buetler, T. Synthesis, tandem MS- and NMR-based characterization, and quantification of the carbon 13-labeled advanced glycation endproduct, 6-N-carboxymethyllysine. *Amino Acids* **2006**, 30, 25–34.

[174] Goldberga, I. , Li, R. , Chow, W. Y. , Reid, D. G. , Bashtanova, U. , Rajan, R. , Puszkarska, A. , Oschkinat, H. , and Duer, M. J. Detection of nucleic acids and other low abundance components in native bone and osteosarcoma extracellular matrix by isotope enrichment and DNP-enhanced NMR. *RSC Advances* **2019**, 9, 26686–26690.

[175] Chow, W. Y. , Bihan, D. , Forman, C. J. , Slatter, D. a. , Reid, D. G. , Wales, D. J. , Farndale, R. W. , and Duer, M. J. Hydroxyproline Ring Pucker Causes Frustration of Helix Parameters in the Collagen Triple Helix. *Scientific Reports* **2015**, 5, 12556.

[176] Chow, W. Y. , Forman, C. J. , Bihan, D. , Puszkarska, A. M. , Rajan, R. , Reid, D. G. , Slatter, D. A. , Colwell, L. J. , Wales, D. J. , Farndale, R. W. , and Duer, M. J. Proline provides site-specific flexibility for in vivo collagen. *Scientific Reports* **2018**, 8, 1–13.

[177] Chung, S. S. M. Contribution of Polyol Pathway to Diabetes-Induced Oxidative Stress. *Journal of the American Society of Nephrology* **2003**, 14, 233S – 236.

[178] Nagai, R. , Hayashi, C. M. , Xia, L. , Takeya, M. , and Horiuchi, S. Identification in Human Atherosclerotic Lesions of GA-pyridine, a Novel Structure Derived from Glycolaldehyde-modified Proteins. *Journal of Biological Chemistry* **2002**, 277, 48905–48912.

[179] Cheng, R. , Feng, Q. , Argirov, O. K. , and Ortwerth, B. J. Structure Elucidation of a Novel Yellow Chromophore from Human Lens Protein. *Journal of Biological Chemistry* **2004**, 279, 45441–45449.

[180] Nakamura, K. , Nakazawa, Y. , and Ienaga, K. Acid-Stable Fluorescent Advanced Glycation End Products: Vesperlysines A, B, and C Are Formed as Crosslinked Products in the Maillard Reaction between Lysine or Proteins with Glucose. *Biochemical and Biophysical Research Communications* **1997**, 232, 227–230.

[181] Nakamura, K. , Hasegawa, T. , Fukunaga, Y. , and Ienaga, K. Crosslines A and B as candidates for the fluorophores in age- and diabetes-related cross-linked proteins, and their diacetates produced by Maillard reaction of α-N-acetyl- <scp>L</scp> -lysine with <scp>D</scp> -glucose. *J. Chem. Soc., Chem. Commun.* **1992**, 992–994.

[182] Shipanova, I. N. , Glomb, M. A. , and Nagaraj, R. H. Protein Modification by Methylglyoxal: Chemical Nature and Synthetic Mechanism of a Major Fluorescent Adduct. *Archives of Biochemistry and Biophysics* **1997**, 344, 29–36.

[183] Sapcariu, S. C. , Kanashova, T. , Weindl, D. , Ghelfi, J. , Dittmar, G. , and Hiller, K. Simultaneous extraction of proteins and metabolites from cells in culture. *MethodsX* **2014**, 1, 74–80.

[184] McHugh, C. , Flott, T. , Schooff, C. , Smiley, Z. , Puskarich, M. , Myers, D. , Younger, J. , Jones, A. , and Stringer, K. Rapid, Reproducible, Quantifiable NMR Metabolomics: Methanol and Methanol: Chloroform Precipitation for Removal of Macromolecules in Serum and Whole Blood. *Metabolites* **2018**, 8, 93.

[185] Delatour, T. , Fenaille, F. , Parisod, V. , Arce Vera, F. , and Buetler, T. Synthesis, tandem MS- and NMR-based characterization, and quantification of the carbon 13-labeled advanced glycation endproduct, 6-N-carboxymethyllysine. *Amino Acids* **2006**, 30, 25–34.

[186] Thompson, J. and Miller, S. N6-(1-carboxyethyl)lysine formation by Streptococcus lactis. Purification, synthesis, and stereochemical structure. *Journal of Biological Chemistry* **1988**, 263, 2064–2069.

[187] Bailey, R. G. , Ames, J. M. , and Mann, J. Identification of new heterocyclic nitrogen compounds from glucose-lysine and xylose-lysine Maillard model systems. *Journal of Agricultural and Food Chemistry* **2000**, 48, 6240–6246.

[188] Ahmed, N. , Argirov, O. K. , Minhas, H. S. , Cordeiro, C. A. A. , and Thornalley, P. J. Assay of advanced glycation endproducts (AGEs): surveying AGEs by chromatographic assay with derivatization by 6-aminoquinolyl-N-hydroxysuccinimidyl-carbamate and application to Nε-carboxymethyl-lysine-and Nε-(1-carboxyethyl) lysine-modified albumin. *Biochem. J.* **2002**, 14, 1–14.

[189] Nagaraj, R. H. , Shipanova, I. N. , and Faust, F. M. Protein Cross-linking by the Maillard Reaction. *Journal of Biological Chemistry* **1996**, 271, 19338–19345.